

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/36598> holds various files of this Leiden University dissertation.

**Author:** Genderen, Eric van

**Title:** Novel detectors and algorithms for electron nano-crystallography

**Issue Date:** 2015-12-03

# **Novel detectors and algorithms for electron nano-crystallography**

**PROEFSCHRIFT**

Ter verkrijging van  
de graad van Doctor aan de Universiteit Leiden,  
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,  
volgens besluit van het College voor Promoties  
te verdedigen op donderdag 3 december 2015  
klokke 13:45 uur

door

**Eric van Genderen**  
geboren te Delft  
in 1981

## Promotiecommissie

Promotores:            prof. dr. Jan Pieter Abrahams  
                              prof. dr. Marin van Heel

Voorzitter:            prof. dr. Jaap Brouwer  
Secretaris:            prof. dr. Mathieu Noteborn  
Overige leden:        prof. dr. Henning Stahlberg  
                              prof. dr. Bram Koster  
                              dr. Jan Visser  
                              dr. Hans Poolman

ISBN-978-94-6259-933-8

© 2015 E. van Genderen

The research described in this thesis was performed at the Division of Biophysical Structural Chemistry, Institute of Chemistry, Leiden University (the Netherlands). This research was funded by NWO, STW, Cyttron I, Cyttron II and Nanomegas®.

This thesis was printed by Ipskamp Drukkers

---

# Contents

<b>Chapter 1</b>	General introduction	<b>7</b>
<b>Chapter 2</b>	Design of cameras based on the Medipix detector family	<b>21</b>
<b>Chapter 3</b>	Protein crystal growth	<b>41</b>
<b>Chapter 4</b>	A Medipix quantum area detector allows rotation electron diffraction data collection from submicrometre three-dimensional protein crystals on Medipix detector family	<b>63</b>
<b>Chapter 5</b>	Lattice filter for processing image data of 3D protein nano-crystals	<b>81</b>
<b>Chapter 6</b>	Ab initio structure determination of nanocrystals of organic pharmaceutical compounds by electron diffraction at room temperature using a Timepix quantum area direct electron detector	<b>95</b>
<b>Chapter 7</b>	General discussion and Future prospects	<b>119</b>
<b>Summery</b>		<b>129</b>
<b>Samenvatting</b>		<b>133</b>
<b>References</b>		<b>136</b>
<b>List of Abbreviations</b>		<b>144</b>
<b>List of Figures</b>		<b>146</b>
<b>List of Tables</b>		<b>147</b>
<b>Publications</b>		<b>148</b>
<b>Curriculum Vitae</b>		<b>149</b>



## About the book cover

The cover of the book was a co-project of José van Genderen-Broers, Ron van Genderen and Eric van Genderen.

The front cover shows a diffraction pattern of multiple summed slices of the carbamazepine dataset used in chapter 6 of this thesis. The pixels of the 512x512 Timepix detector are clearly visible. The image has been enhanced with a blue color.

The back cover of the thesis supports an early cover design as was drawn by RvG. I really wanted this early design to be used, because it symbolizes the work from early drawing, to final design in method development.

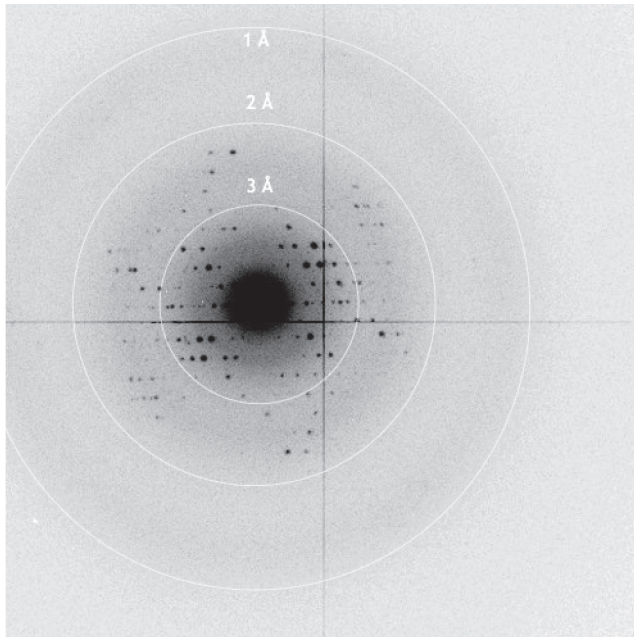
**It doesn't matter how beautiful your theory is,  
it doesn't matter how smart you are.  
If it doesn't agree with experiment, it's wrong**

**Richard P. Feynman (Physicist, 1918-1988 )**



# Chapter 1

## General introduction



## Contents

<b>1.1 Introduction .....</b>	<b>9</b>
<b>1.2 Transmission Electron Microscopes.....</b>	<b>10</b>
<b>1.3 Electron scattering and radiation damage.....</b>	<b>12</b>
1.3.1 Imaging versus diffraction .....	14
<b>1.4 Organic molecules .....</b>	<b>15</b>
1.4.1 Small and big organic compounds .....	16
1.4.2 Proteins .....	17
1.4.3 Single particle imaging .....	18
<b>1.5 Crystal diffraction.....</b>	<b>18</b>
1.5.1 2D crystals by TEM .....	18
1.5.2 3D crystals by TEM .....	19
1.5.3 XFEL another new technique .....	19
<b>1.6 Thesis outline .....</b>	<b>19</b>

**Chapter image:** this is a very nice example of a diffraction pattern; the almost perfect symmetry reminds of stars, the spots as well as the shape (Fig 1.4)

---

---

### 1.1 Introduction

Electron Diffraction is used for structure determination since early last century. The first years it was mainly used to study metals, alloys, etc. The different techniques which have been developed for either imaging or diffraction are numerous, for example: alloy determination, consistencies of metals, glasses and polymers. (Williams and Carter, 2009). However, to study samples of organic biological origin soaking them in heavy metals, replacing the water (dark staining) was the only way to image them; by using heavy metals the contrast could significantly be improved. This method is still being used a lot, to study tissues in biology. It is, however, almost impossible to use heavy metal soaking to determine the structure of smaller things like proteins, or to achieve a resolution better than 12-20Å. Fifty years ago, a new technique was developed for Electron Microscopy (EM) which involved cryo-freezing the sample to -180 °C by plunge freezing (Taylor and Glaeser, 1974; Klug, 1979). This made it possible to study biological and organic samples embedded in amorphous ice for a longer time and at higher energies. Because of evolving technologies and the introduction of direct electron detectors like the Gatan K2 (Gatan, 2015) and the FEI Falcon 2 cameras (FEI 2015), resolutions of up to 2.2Å are now reported (Bartesaghi et al., 2015).

## 1.2 Transmission Electron Microscopes

A big variety of transmission electron microscopes (TEMs) is available (for example: HRTEMs, STEMs, HVEMs and cryo-TEMs), while all are slightly different, they are based on the same basic TEM principle: Using electrons which are transmitted and scattered through the sample and can then visualized by different electron detection systems.

Currently two different techniques exist to generate electrons in TEM with an Electron Gun (EG): thermionic and field emission guns. Thermionic sources consist of either tungsten filaments or lanthanum hexaboride LaB<sub>6</sub> crystals. The latter is now more common than the former. The Field Emission Gun (FEG) is getting more common as a source of electrons in TEM, especially at high end machines. The beam can be smaller, more monochromatic and more coherent than the older thermionic EGs. This improves the Signal to Noise (S/N) ratio and spatial resolution.

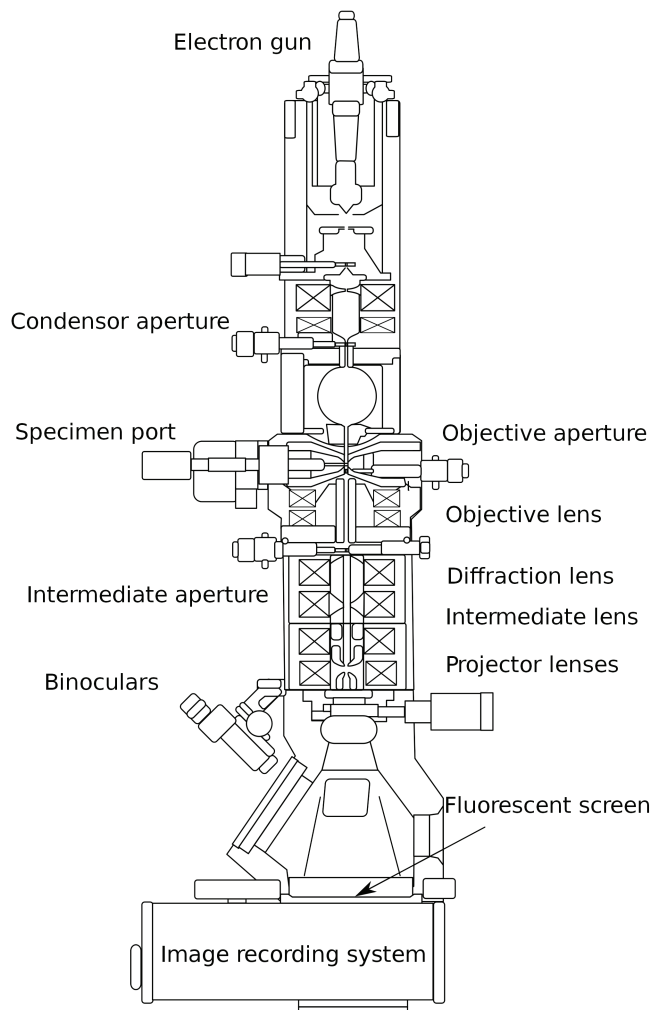
The illumination area of the sample is determined by one or more condenser apertures and lenses. The condenser apertures limit the amount of electrons which move through the illumination system towards the sample. The condenser lenses are also responsible for the exposure area on the sample, by changing to a converged, parallel or divergent beam. In a parallel beam condition the condenser aperture determines the total exposed area on the sample.

The beam passes through the objective aperture, the sample, the objective-, diffraction-, intermediate and projector lens as well as the selected area aperture. After which it enters the camera chamber.

On older TEMs the image is usually formed on the fluorescent screen which can be observed through a lead-glass window. On more modern TEMs this fluorescent screen is no longer visible from the outside and is integrated in the TEM with a (CCD) camera. Below the fluorescent screen is usually room for one or multiple detectors and/or imaging systems. Originally these detection systems consisted either of film, image plate or CCD cameras. (Figure 1.1 shows the common lay-out of a TEM).

Recently direct electron detectors have been developed like the Gatan K2 (Gatan, 2015) and the Falcon 2 (FEI, 2015), which have, for imaging, a much higher throughput and S/N ratio than CCD cameras (McMullan et al., 2014). Because of the way they are designed these cameras are easily damaged when used during diffraction experiments. To prevent damage to these sensitive

cameras, hard coded restrictions are in place, to make sure they are not used under such locally intense beam conditions. Because of this, they are not suitable replacements for either image plate, film or CCD cameras, when used for diffraction experiments. An example of a detector which can handle these more extreme condition is of the Medipix family (Chapter 2). The full energy of the impacting electron is captured in the sensitive layer and the underlaying electronics are thereby protected from damage. Also the ability to discriminate between high energy electrons and particles with a lower energy (for example the abundant 5-10 keV X-rays) make it possible to detect Bragg spots with intensities of only a couple of electrons above the electron scatter noise.



**Figure 1.1: Typical transmission electron microscope schematic.** Gun system, apertures and lens systems. (Source wikipedia)



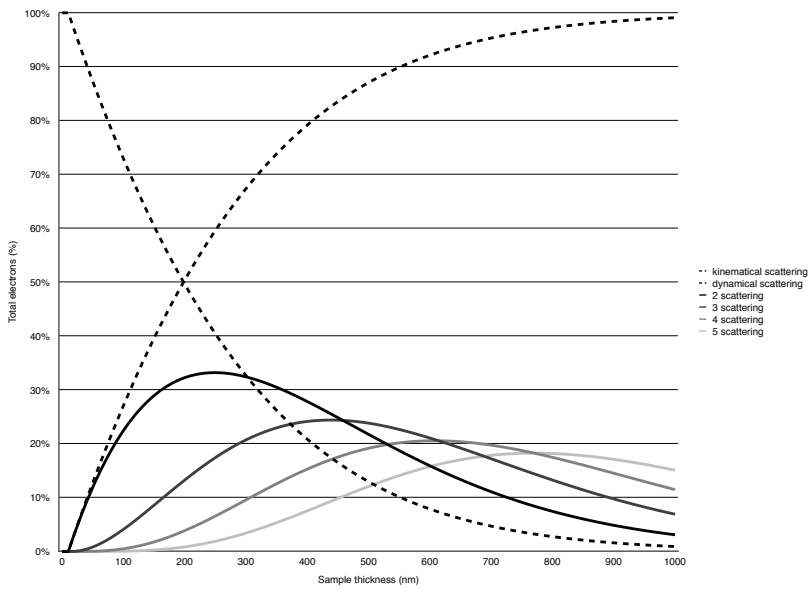
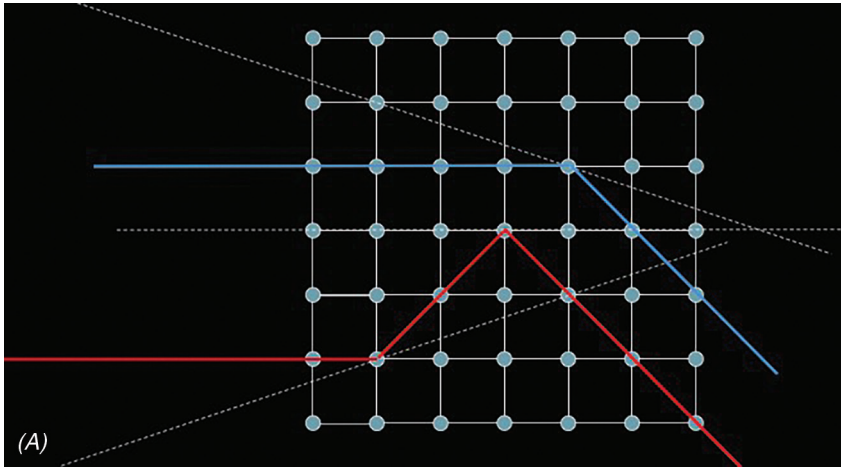
## 1.3 Electron scattering and radiation damage

An electron is a negative charged particle with low mass, therefore it is easily influenced by other electrons and/or the positive nucleus of an atom. These Coulomb interactions are well described in literature and understood. These interactions cause the electron to scatter, either elastic, inelastic or dynamically. (Wang, 1995; Cheng, 2013; Williams and Carter, 2009)

Elastically scattered electrons are usually coherent, which means that they stay in phase with each other. Elastic scattering usually occurs at low angles of about  $1-10^\circ$ , where at higher angles they become more and more incoherent. Inelastic scattering is in almost all cases incoherent and scattering usually occurs under  $< 1^\circ$ . Dynamical scattering happens when an elastically scattering electron scatters multiple times in the sample. The thicker the sample the more often this can occur, this makes it almost impossible to determine where the electron scattered the first time. (Fig. 1.2). Lastly there are the electrons which are unaffected when transmitted through the sample (end up in the central beam in diffraction mode). As soon as the electrons have been transmitted through the sample the electrons are no longer in phase final which results is in a non uniform distribution.

**Table 1.1: Energy deposited in biological/organic specimens per useful scattering event.**  
(Henderson, 1995)

	Electrons	X-rays	
	200-300 keV	1.5Å	30Å
<b>Ratio (inelastic/elastic)</b>	3	10	~5000
<b>Mechanism of radiation damage</b>	Secondary e <sup>-</sup> emission	Photoelectric e <sup>-</sup> emission	
<b>Energy deposited per inelastic event (eV)</b>	20	8000	400
<b>Energy deposited per elastic event</b>	60	80'000	400'000
<b>Energy deposited relative to electrons</b>			
(inelastic)	1	400	20
(elastic)	1	1000	10000



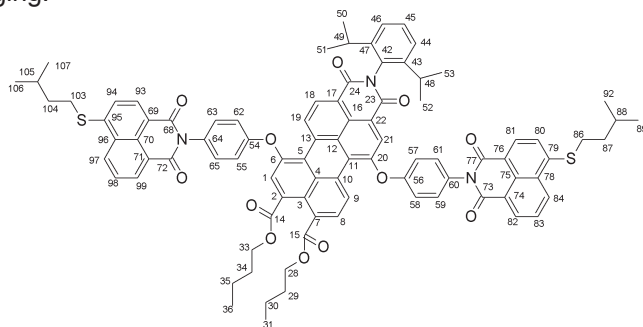
**Figure 1.2: Scattering in crystals.** Top: kinematic 'single' diffraction (blue) coincides with dynamic diffraction (red). Bottom: fraction of scattered electrons in protein crystals at 300keV

### 1.3.1 Imaging versus diffraction

In imaging the spatial distribution of the electrons scattering is used; this can be observed as contrast in the image. This contrast can be enhanced when using the objective aperture. When doing diffraction experiments the angular electron scattering distribution is used. In imaging the unaffected electrons are interfering with the information the scattering electrons carry by adding shot noise, whereas for diffraction all these non-information carrying electrons are contained in the central peak (direct beam). Therefore, in diffraction, images are considered to be less 'noisy', because there is no shot noise in added to the information carrying signal.

Inelastic scattering leads to beam damage of the sample. This damage is called radiolysis: the interaction of the electron transmitted from the EG with usually another electron in the sample (e.g. ionisation). Radiolysis breaks the chemical bonds between atoms within a molecule. Two more types of damage also exist: (i) Sample heating by phonons. This is a major source of damage in organic materials (hence cryo-cooling). (ii) Knock-on damage, where the atom gets displaced in the lattice. Because this effect is more rare than the other two the knock-on effect is much less a factor when working with beam sensitive materials like organic compounds, proteins or zeolites.

The most common way to prevent beam damage is to: (i) use techniques which make use of a low electron dose on the sample. (ii) coat the specimen with heavy metals; (iii) to cool the sample to very low temperatures. Low temperature does not only prevent beam damage but also prevents the sample from losing all the water content to the vacuum of the TEM when the sample contains water. (Protein crystals contain 35%-70% water). While the beam damage described above is a problem in EM, X-rays tend to damage the sample more per diffracted quantum than electrons (Henderson, 1995) (Table 1.1). Thus the interest in researching new EM methods for protein 3D protein crystallography and single particle imaging.



**Figure 1.3: Chemical structure perylene derivative.**

---

---

## 1.4 Organic molecules

The size of organic molecules can vary from a few atoms to thousands of atoms in big structures like proteins. These molecules can be studied by different techniques like X-ray crystallography, SSNMR and single particle cryo-EM for big protein molecules and clusters. The importance of solving the 3D structure can vary by the type of molecule as well as its function. For example, the crystal structure information or packing is more important than solving the 3D structure for small organic compounds; because the crystal size and structure can say something about properties like solubility, which is very important for drugs. Secondly for proteins the packing itself is often not interesting but is essential for solving the 3D structure. Knowledge the 3D structure will help biochemists to understand the function and active sites of the protein, which may help to develop the drugs.

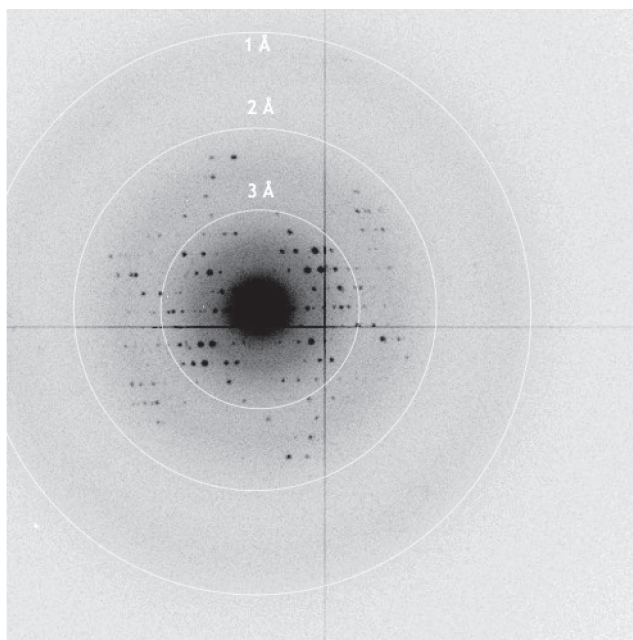


Figure 1.4: Diffraction pattern from perylene derivative.

### 1.4.1 Small and big organic compounds

There are many different techniques available to look at the packing of small organic crystals, for example: (i) single crystal X-ray diffraction and solid state NMR (SSNMR) for micrometer sized crystals. (ii) X-ray / electron powder diffraction for smaller crystals.

When these organic crystals are used in drugs or food additives, one of the most important characteristics is their solubility; this is directly linked to how the compounds are stacked in the crystal (Ozaki et al., 2014). While crystals can usually grow into bigger crystals, smaller crystals in the same sample can be a totally different crystal polymorph. For example, in powder diffraction it is possible that the signal of a low-abundant polymorph (< 5%) can be completely hidden within the brighter signal of the more abundant polymorph. (Chapter 6)

Because of technical advances in EM, many different methods have been developed for looking at 3D nano-crystals of organics with Cryo-TEM. Examples are: automated diffraction tomography (ADT) (Kolb et al., 2008; Kolb et al., 2007; Mugnaioli et al., 2009) and rotation electron diffraction (RED) (Wan et al., 2013; Yun et al., 2015; Zhang et al., 2010). These two techniques show solved crystal structures of beam sensitive compounds *ab initio* from 3D ED data acquired at low electron dose from cryo-cooled samples. However, there are multiple reasons to consider doing experiments at ambient temperature when possible. (i) Interference of ice crystals by ice formation during sample load; (ii) necessity of in-column cryo-plates to reduce contamination; (iii) possible change in crystal structure; (iv) reduced sample throughput, as loading/unloading samples can take up to 2 hours. There are examples of beam sensitive materials solved at ambient temperature by ED, but these structures were solved using the non-continuous ADT technique, e.g. Benzamides (Gorelik et al., 2012) and 9,9'-bianthracene-10-carbonitrile (CNBA) (Kolb et al., 2010).

---

In chapter 6 we discuss a method which overcomes these challenges by solving the structure of compounds *ab initio* at ambient temperature conditions of sub-micrometer nano-crystals. We demonstrate this technique with two pharmaceutical organic compounds (carbamazepine and nicotinic acid). Data from XDS were prepared with XPREP (Bruker, 2004) for *ab initio* structure solution using the program SHELXT; subsequent refinement of the given structure was done with SHELXL and ShelXle (Sheldrick, 2008; Hübschle et al., 2011). The eight parameter fitting for electron scattering factors from (Peng, 1999) was used for SHELXL. The carbamazepine structure was refined anisotropically unrestrained except for the RIGU restraint (Thorn et al., 2012). Nicotinic acid was refined isotropically without any restraints.

The method has also been used to get an estimate of the packing of a perylene derivative crystal, a large antenna complex of 1427 g/mol. The crystals of this compound do not grow into consistent crystals of sufficient size to be analyzed by X-ray crystallography. TEM electron nano-diffraction allowed to estimate the unit cell with XDS, which than could be used to constrain the data obtained by SSNMR and X-ray powder diffraction (XRPD). This shows that a combination of these techniques can be combined. (Not discussed in this thesis. Added for completeness)

### 1.4.2 Proteins

There is a huge interest in the 3D structure of proteins. The structure can give information about the purpose but also malfunctions of the protein when certain DNA mutations inserted a different amino acid in the chain. Also the 3D structure makes it possible to design drugs or test compounds *in silico*. Because of this interest in the 3D structure different methods and techniques have been developed not only in X-ray but also in EM. The main techniques for EM are described below.

### 1.4.3 Single particle imaging

Currently, single particle imaging of large protein complexes is evolving at a rapid pace. Recently a structure of  $\beta$ -galactosidase has been reported with a resolution up to 2.2Å (Bartesagh et al., 2015). One of the reasons is the quick advancement in resolution comes with the developments in semi conductor technology. This made it possible to design direct electron detectors like the Falcon 2 and the K2. Also the increase in computing power in combination with new algorithms make it possible to rapidly analyze big datasets which consist of thousands of images with hundreds of thousands particles in random orientations. However, also the improvements are made in the microscopes themselves with Cs aberration correctors (Batson et al., 2002), better coherent XFEG guns and phase plates (Marko and Danev, 2012).

## 1.5 Crystal diffraction

Another way of solving 3D structures of proteins is by crystallizing them. There are however some bottlenecks when it comes to solving protein 3D structures from crystals. The most important of these is that some proteins do not crystallize into the micro size crystal that is needed for X-ray diffraction systems. If it is too small the crystal will be destroyed before sufficient data can be acquired.

### 1.5.1 2D crystals by TEM

2D crystals are often used for membrane proteins and usually contain lipids and detergents. Techniques available from single particle imaging can be used since the crystals are made out of a single layer of proteins which are aligned all in the same orientation. However 2D crystals are prone to bending and cracking. Algorithms can “unbend” the crystal. The sample is usually rotated and a tomography series is taken of the crystal. Because of the high symmetry, diffraction patterns can be taken to help with the unbending and padding of the crystal. The crystal is usually vitrified or negatively stained. (Henderson and Unwin, 1975; Gil et al., 2006; Zeng et al., 2012; Raunser and Walz, 2009)

---

### 1.5.2 3D crystals by TEM

Large protein(s) (complexes) can be solved with single particle imaging and micrometer sized 'thick' crystals can be solved with X-ray diffraction; however, some proteins are just too small to be solved with single particle EM and other proteins do not grow into the large crystals needed for X-ray crystallography. These proteins form a target for electron nano-crystallography with TEM and X-ray Free Electron Laser (XFEL). Currently two approaches are developed for TEM: the microED method (Hattne et al., 2015) and our method of thin 3D electron nano-crystallography (Chapter 4, 5 and 6).

### 1.5.3 XFEL another new technique

X-ray Free Electron Lasers (XFELs) are also a valid instrument to study 3D nano-crystals. These can be used to collect datasets from crystals as small as 150 nm. Many crystals are needed to collect one full dataset since every crystal can only be measured (1-5 ml highly concentrated crystal solution is needed). With the intensities used in the XFEL the crystal gets vaporised (Lomb et al., 2011). After collection of the full dataset the pseudo randomly orientated diffraction datasets need to be analyzed and at least 10k images are needed for high resolution 3D reconstruction (comparable to single particle imaging). (Kiran et al., 2011)

Because it is really hard to get a good sample for XFEL and the running costs an interesting solution to check the sample quality is to diffract them on TEM before acquiring a dataset on the XFEL. The XFEL allows collecting data at room temperature and recently also previously unknown structures have been solved using XFEL data (Kang et al., 2015)

## 1.6 Thesis outline

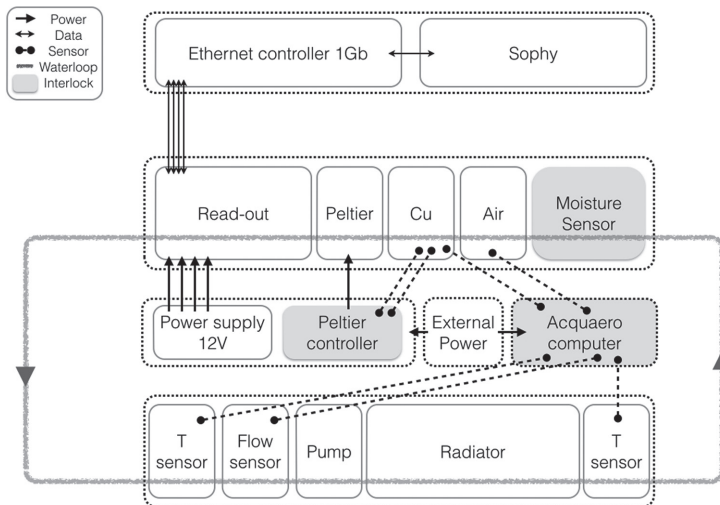
Combining the phases acquired from images of protein 3D nano-crystals (Nederlof et al., 2013a) and diffraction data from electron nano-crystallography should make it possible to solve the structure of proteins. In this thesis I describe: (i) use of the special Timepix detector to get excellent diffraction data (Chapter 2, 4 & 6); (ii) growth of protein crystals (Chapter 3); (iii) acquiring diffraction datasets with electron microscopy (Chapter 4 & 6); (iv) improvement on the algorithm and filtering of images of 3D protein crystals (Chapter 5). As well as using these techniques for *ab initio* structure determination of organic compounds by direct methods without the use of cryo-TEM (Chapter 6).





## Chapter 2

# Design of cameras for TEM systems based on Medipix detector family



## Contents

<b>2.1 Introduction .....</b>	<b>23</b>
<b>2.2 Chip, readout and software .....</b>	<b>23</b>
2.2.1 Chip versions.....	24
2.2.2 Chip design .....	25
2.2.3 Silicon layer characteristics, scattering and data quality .....	26
2.2.4 DQE.....	28
2.2.5 Timepix ASIC.....	29
2.2.6 Read-out.....	30
2.2.7 Software .....	31
<b>2.3 Camera design .....</b>	<b>32</b>
2.3.1 Cooling .....	32
2.3.2 YEOL 1010 Medipix camera .....	32
2.3.2.1 Chip & read-out modifications.....	33
2.3.2.2 Software.....	33
2.3.3 Quad Timepix for a Titan Krios “Hardy” .....	34
2.3.3.1 Chip & read-out.....	34
2.3.3.2 Pod design.....	34
2.3.3.3 Cooling.....	36
2.3.3.4 Safety.....	36
2.3.3.5 Software and Hardware .....	36
2.3.4 Single quad Timex design “Laurel”.....	37
2.4.3.1 Pod design.....	37
2.4.3.2 Cooling.....	37
<b>2.4 Discussion .....</b>	<b>38</b>

**Chapter image:** design schematics show in a simple way the intrigate design descisions that were made, to build a camera that is safe to use. (Fig 2.3)

---

---

## 2.1 Introduction

The detector is one of the most important focus points for the development of electron nano-crystallography, as the method has been developed in the past 8 years. During this time many different detectors and housings have been designed to be used in different transmission electron microscopes (TEMs). The past five years this effort has been intensified to design better and more robust housings for longer lasting experiments; minimizing the chance for damage of the electronics and optimizing the overall stability. In this chapter I describe the basic technology behind the Medipix-type Timepix detector and how it behaves in TEM and I explain how these detectors can be build into TEMs by using three examples of the most important camera iterations (the preconditions, solutions, possible problems and bottlenecks).

## 2.2 Chip, readout and software

Late in the 1990s the first Medipix collaboration was formed at CERN in cooperation with Nikhef. The goal was to design a functional hybrid detector to detect quanta of all types and energies with medical applications in mind, hence the name Medipix. A decade later the Medipix 2 consortium was looking for a successor. Three main types of Medipix-2 have been designed and produced so far: Medipix-2, Medipix-2 MXR and the Timepix (note the number abbreviation). Currently the Medipix 3 collaboration is in full swing and has designed the Medipix-3 (RX) and Timepix-3 chip. At the moment of writing the Medipix-4 collaboration is being formed. A short overview of the chips and their characteristics (Source: Wikipedia & Medipix collaboration website).

## 2.2.1 Chip versions

- Medipix-1: 64x64 pixels (4k) with a 170  $\mu\text{m}$  pixel pitch. It had a speed of 2 MHz per pixel with a counting depth of 15-bit. (Bisogni et al., 1998)
- Medipix-2: 256x256 pixels (65.5k) with a 55  $\mu\text{m}$  pixel pitch. Instead of the Medipix-1 it had a upper and lower energy threshold, the maximum count rate was 100kHz per pixel. (Llopart et al., 2002)
- Medipix-2 MXR: Successor of the Original Medipix-2 chip. It has a better temperature control, stability and pixel overflow. Also it has an increased radiation hardness. (Medipix, CERN)
- Timepix: By dropping the upper threshold from the Medipix-2 it has the ability to use a Time Over Threshold (TOT) and Time Of Arrival (TOA) mode. The amount of energy is being counted by the TOT mode and the TOA mode measures the time between a trigger pulse and the arrival of the charge into each pixel. (Llopart et al., 2007)
- Medipix-3: This has the same pixel lay-out as the Medipix-2 chip however when binned to pixels of 110 micrometer it has 8 different energy thresholds as well as 4 thresholds with continuous read-out. Also improvements have been made in the energy resolution by having a real time charge sharing correction. (Ballabriga et al., 2011)
- Timepix-3: The spiritual successor of the Timepix (note: there is no Timepix-2). Where all other chips are having a frame based read-out, the Timepix-3 has an event driven read-out system. Therefore as soon as a pixel gets triggered above the set threshold the signal will be recorded along with the location.

**Table 2.1: Comparison table between currently available chips.**

	Medipix2	Timepix	Medipix3	Timepix3	Medipix4	Timepix4
Pixel side ( $\mu\text{m}$ )	55	55	55/110	55	x/2x/3x	y
Technology (nm)	250	250	130	130	65	65
# pixels in x and y	256	256	256/128	256	512/256/128	512
Readout architecture	Frame based Sequential RW	Frame based Sequential RW	Frame based Continuous RW	Data driven/ frame based	Frame based Continuous RW	Data driven/ frame based
Charge summing and allocation mode (CSM)	No	No	Yes	No	Yes	No
# thresholds	2 (window discriminator)	1	2/4/8 Seq RW 1/4 Cont RW	1	?	1
ToT/ToA	No	ToT (14 bit) OR ToA (14 bit, 10ns precision)	No	ToT (10 bit) AND ToA (14 bit, 1.56ns precision)	No	ToT AND ToA
Front end noise ( $e^-$ rms)	110	100	80(SPM) 174(CSM)	62	$\leq 80$ (SPM) $\leq 174$ (CSM)	$\leq 62$
Peaking time (ns)	150	100	120	30	$\ll 120$	$\ll 30$
Max count rate ( $\text{Mc}/\text{mm}^2/\text{s}$ )*	826	-	826 (SPM 55 $\mu\text{m}$ ) 164 (CSM 55 $\mu\text{m}$ ) 376 (SPM 110 $\mu\text{m}$ ) 28 (CSM 110 $\mu\text{m}$ )	0.43 (data driven)	x5 Medipix3	x10 Timepix3
Number of sides available for tiling	3	3	3	3	4	4

\*Depends strongly on exact conditions of threshold, sensor material and energy of illumination  
Brown indicates parameters which are still to be defined

---

In this thesis both Medipix-2 (chapter 4) and Timepix (chapter 6) detectors were used to measure data. However, all detectors that are still actively used are of the Timepix type. The reason is that the Timepix has an increased stability over the Medipix-2. Also, there is no need for an upper threshold because the highest energy particles in a TEM are the electrons emitted from the gun. Since the Medipix-2 is no longer in active use and therefore only the Timepix will be discussed in depth in this chapter.

### 2.2.2 Chip design

The Timepix chip is a hybrid detector (A semi-conducting material is bump-bonded to an electronics layer). In our case we only used a semi-conductor made out of a single Si crystal, but others types like GaAs and CdTe do exist. The latter two have different characteristics. Because of the higher density (Z value) of the detector layer they will make the electrons scatter more locally, this will increase the spatial resolution of the data. However there are drawbacks: (i) more electrons will backscatter and get “lost”; (ii) require more equalization, because of irregularities in their crystal structure; (iii) more difficult and therefore more expensive to be produced. (Zwerger et al., 2007; Hamann et al., 2015; Rasif, 2015).

When a high energy particle (in TEM mostly electrons and X-rays) scatters through the sensor layer, an electron/hole pair cloud is created. By putting a bias over the sensor layer, the positive (or negative, depending on the type semi conductor) charge cloud is conducted to the ASIC (application-specific integrated circuit) electronics. In the analogue part of the electronics the deposited energy is amplified and, if the total energy is higher than a calibrated threshold energy setting (per pixel), the energy peak will result in counts in the digital part of the electronics. (Fig. 2.1).

The Timepix offers three different counting modes: (i) Standard counting mode (CM). If the signal in a pixel is above the threshold the counter will increase by one (independent of how much charge is deposited above the set energy threshold). (ii) Time Over Threshold (TOT). If the energy deposited in the pixel electronics is higher than the energy threshold the capacitor is slowly depleted. During this depletion the amount of clock cycles is counted. This gives a measure of the amount of energy deposited (Fig. 2.1). (iii) Time Of Arrival (TOA). This mode is, for now, not very useful in EM. If the TOA is triggered a time stamp is being made.

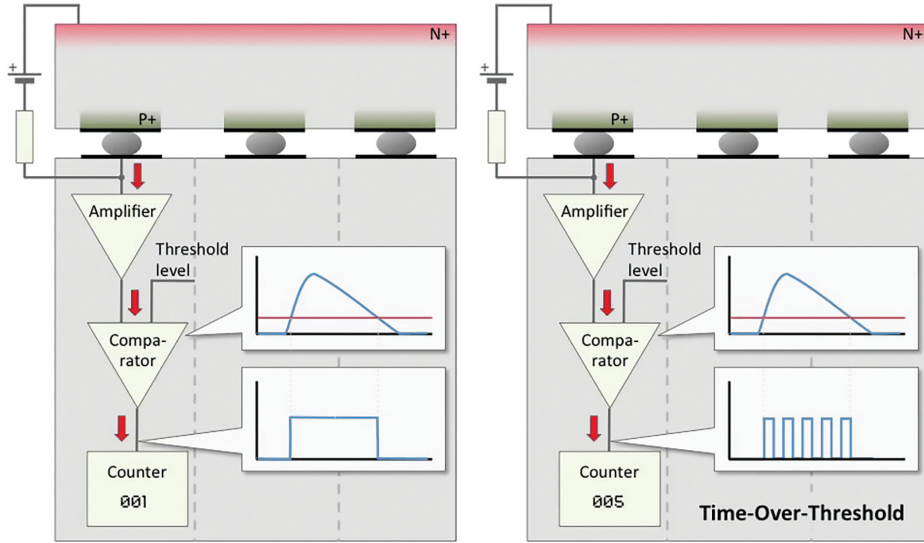
In the work described in chapter 4 & 6, usually a combination is used of four single chips together bonded with a single Silicon layer: a Medipix or Timepix Quad.

---

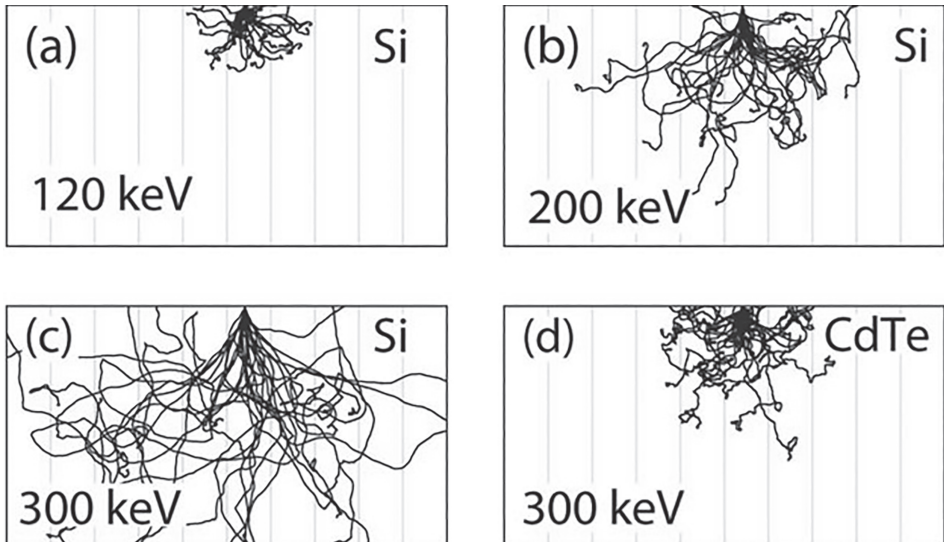
### **2.2.3 Silicon layer characteristics, scattering and data quality**

When electrons go through the sensitive Silicon layer they will start scattering. While scattering through the Silicon the electron will gradually lose energy by generation of Bremsstrahlung (X-ray). Scattering is the main contributor to the generation of electron-hole pairs. If the incident electron has a higher energy, it will scatter further through the Silicon. Therefore using lower energies for TEM will result in better counting statistics, since the chance that the electron will also deposit energy in neighboring pixels will be reduced. Early in the track through the silicon the angles at which the electron scatters are usually low because of forbidden angles. When the electron loses more energy scattering events will occur more regularly and the angles at which the electron scatter can scatter are greater. This explains why most of the energy deposition is localized late in its track through the silicon. The tracks of these electrons through the silicon layer have been thoroughly simulated and tested for the Medipix-2 at different energies (McMullan et al., 2007; McMullan et al., 2009; Faruqi & McMullan, 2011) (Fig. 2.2).

It is possible to use the scattering at high energies in silicon as an advantage by using the Timepix-3. It could allow us to track the scattering paths of the electrons. The timing, the amount of charge deposited in multiple pixels as well as the shape of the charge cloud will all be required for determining the position of impact. The sharpness of the image could improve significantly. This would only work if every impact can be considered separate (high speed read-out and low dose would be needed). Early tests and simulations show that this technique is promising, but more experiments are required and therefore it is not described in this thesis.



**Figure 2.1: Counting modes in Timepix chip.** (Left) Once the charge cloud from the incident quanta reaches the ASIC, the signal is amplified and if the signal is over a certain threshold the counter is increased by one. (Right) Once the charge cloud from the incident electron reaches the ASIC, the signal is amplified and if the signal is over a certain threshold the clock cycles are counted till the capacitor is depleted and its energy level is below the threshold. (Source: ASI, 2015)



**Figure 2.2: Monte Carlo simulations of electron scattering.** Different materials and at different energies. The horizontal lines simulate pixels of 55 micrometer wide. (Faruqi & McMullan 2011)



## 2.2.4 DQE

The Data Quantum Efficiency (DQE) of Medipix type chips with a Silicon sensor layer is excellent for energies of 120 keV and lower. Higher energies than 120 keV will result in lower DQE (Mcmullan et al., 2007; McMullan et al., 2009). This is the result of the larger distance over which electrons at higher energies loose their energy. This can lead to counts in other pixels than the one of impact, especially if the energy threshold is set very low. If the threshold is set to a higher energy the DQE can be improved. However, the change increases where the electron is not counted at all (because of an electron scattering over multiple pixels and in each pixel depositing an energy which is too low to be accepted). Therefore, when setting the energy threshold (THL) level, one must compromise between having a chance of charge sharing or losing electrons. This effect can be reduced by using a more dense (higher Z) semi-conductor as sensor, resulting in a better DQE for electrons at higher energies.

There is a quantitative way to describe the quality of detector performance (DQE). The DQE is a combination of the Modulation Transfer Function (MTF) and the Noise Power Spectrum (NPS). The MTF describes how much of the contrast information can be transferred from the object to the image at each resolution. The MTF is a Fourier transform of the Point Spread Function (PSF).

$$\mathfrak{F}(\text{PSF}) \equiv \text{MTF}$$

DQE is the Signal to Noise Ratio (SNR) of the signal reported by the detector divided by the SNR of the signal. The relation between DQE and MTF/NPS can be found in equation 2.1.

$$\text{DQE}(\omega) = \frac{n_{out}^2 * \text{MTF}(\omega)}{n_{in} * \text{NPS}(\omega)}$$

**Eq. 2.1:** ( $n_{out}$  is the average signal of output image,  $n_{in}$  is input electronic dose,  $\omega$  is the single spatial frequency variable (Shaw, 1978))

This formula shows why the DQE of a Medipix-2 type detector degrades at higher energies and lower THL: electrons are counted in wrong pixels or double counted in two neighboring pixels and the incoming electron dose is no longer in line with the recorded image. This lowers the MTF and increases the NPS. By setting the THL high it is possible to mask the double counts while losing electrons (this influences the DQE by lowering the NPS but increasing the input electron dose to the average signal of the output). However the Medipix still has a good SNR because of having a negligible (electronic) read out noise, which improves the DQE significantly over standard CCD cameras, which have dark-current and read-out noise. (Mcmullen et al., 2009)

---

In low dose electron nano-diffraction, the DQE alone is insufficient as an estimator of data quality. The following properties of a detector describe the quality of a detector for diffraction better than the DQE: (i) High dynamic and linear range of multiple incident electrons in a pixel. (ii) Exact count of the deposited electrons in a Bragg spot compared to other Bragg spots. (iii) Having almost no read-out noise to catch even the faintest Bragg spots of up to three electrons above the electron background noise. (iv) High contrast between very bright Bragg spots and nearby very weak Bragg spots. (v) Relative radiation hardness of the detector. (For examples see chapter 6). All of these properties are combined in the Medipix type detectors. The THL should be chosen in such a way that fits the experiment (e.g. If the THL is chosen low, the data quality will deteriorate slightly, because double counts will be recorded).

### 2.2.5 Timepix ASIC

Electronics of one single pixel can be described as a single very small 1D detector and is of a complicated design because of the space constraints in combination with the used semi-conductor technology. A simple scheme shows the analogue and digital electronics of a single pixel (Fig. 2.3). After the energy has been deposited in the silicon the energy will be (in the form of electron-hole pairs), under influence of a bias be multiplied in an amplifier of the ASIC and then be send to the Discriminator. Here the Threshold (THL) setting for this chip is compared to the incoming energy and if necessary the Threshold output is corrected for each pixel individually with a maximum value of 4 bits. Then, if a pixel is supposed to be masked (because of being dead, bright or unstable) the counter will be overwritten and a 'zero' is communicated. From this point the information enters the digital part of the ASIC. At this point the reference clock is introduced as well as the shutter. The signals are then synchronized over the other pixels in the column and the read-out clock (frame time) starts to collect the amount of counts detected in this frame which is then send of to the 14 bit register to be added to the 'line' read-out.

## 2.2.6 Read-out

The read-out system is the hub between the computer/software and the chip. It is a two way gated transportation hub for Data Acquisition Commands (DACs) towards the chip and frame read out towards the computer. The read-out system is usually also responsible for supplying the bias voltage of the Silicon layer as well as the power for the chip.

Two different read-out systems have been used in our projects. For the Medipix-2 we used a USB1.1 (Vykyda et al., 2006) and the FITPix read-out system with a USB connector (Kraus et al., 2011). Since the chips were of older Medipix versions, continuous read-write was not possible. Also the read-out was not able to supply enough power for the total power consumption of a quad. Therefore, the detector (Chip & Read-out) only supported 3.3 fps with a maximum 50% dead time (DT) per second. This combination was used in the chapter about rotational crystallography of protein crystals (chapter 4).

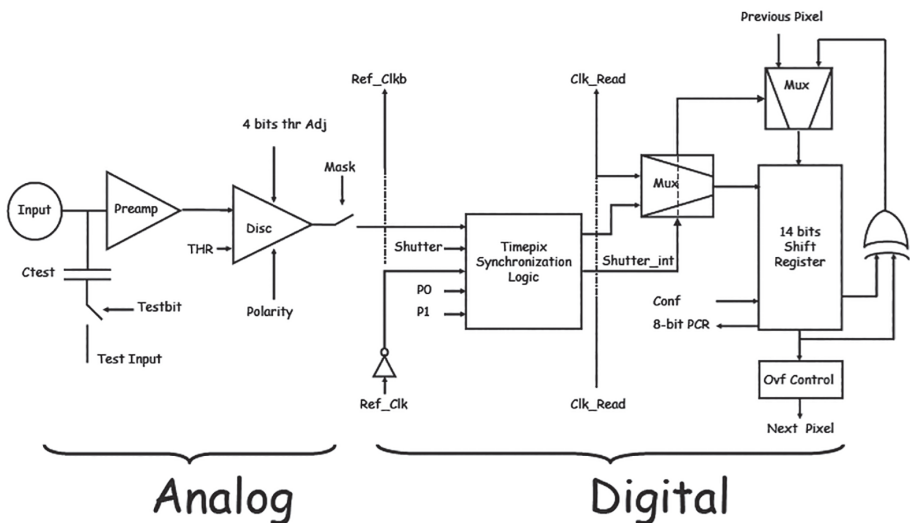
For the Timepix chip we used the RELAXD board, designed by Nikhef (Visser et al., 2011). This read-out board utilizes the full potential of a 1Gb ethernet connection. Therefore, it supports a maximum of 120 fps for a single Timepix Quad with a DT of 8.3 ms per frame. For the experiments with the Timepix we usually set the exposure per frame to 100 ms, which resulted in a total 9.1 fps.

## 2.2.7 Software

Control of data acquisition requires sending instructions to the chip and getting the frame information back. The instruction set for the Medipix-2 and Timepix are well documented within the collaboration, and therefore many basic instruction sets exist. However, working with them requires advanced knowledge of command-line shells. For our work we used two software packages with a user interface: Pixelman (Turecek et al., 2011) and SoPhy (ASI, 2015). To operate the detector using USB read-out boards, Pixelman is the best solution and SoPhy is optimized to be used with the Timepix chip and Relaxd read-out board. Both these software packages support alignment scripts for: (i) chip equalization, where the different pixels are equalized based on the noise edge of the chip.;(ii) dead, bright and noisy pixel detection.

The scripts make the tedious work of alignment much easier. Since the noise edge is far away from the operating THL value, the chip to chip alignment still requires manual adjustments. Sophy is able to align 16 or more individual chips.

While these software packages are mainly designed for alignment and chip testing, they have been further developed for X-ray imaging and diffraction and in the case of SoPhy also Mass Spectrometry. Currently (2015) we only use the SoPhy software package. In the near future this package will contain a module specially designed for ED, which removes many small and inconvenient user problems (e.g. live image and data acquisition at two different fps). Sophy is able to store images in .txt, .tif and .bin formats.



**Figure 2.3: Simplified schematic of the counting electronics of a Timepix chip.** (Source: Medipix consortium)

## 2.3 Camera design

### 2.3.1 Cooling

The first camera systems designed for the Medipix-2 v1.3 detector were not actively cooled (Georgieva et al., 2011) (chapter 4). However, we found that not cooling the detectors or read-out could lead to a problem: when the temperature is not stable, several DACs could become unstable (Lower Threshold: THL and THS mainly). The main problem is that the different chips of a quad detector are equalized for a certain threshold energy. Chips do not have the same behavior to shifts in temperature. Therefore if a detector is heated during operation and impact of electrons, the energy calibrated to a set THL will also slightly change for the different chips. At 50-60 °C also an increase in electronic noise becomes noticeable. Because of these two effects in new camera iterations the cooling systems were improved with the aim to keep a stable temperature.

The first iteration of the camera series supported a Medipix on a Quad Nikhef carrier board and a Fitpix (Vykydal et al., 2006) read-out. In an early designs this detector was only passively cooled to the outside air by a copper bridge, this was improved in the second version by actively cooling the detector to a CPU water-cooling system at RT. Early results made clear that not only the detector but also the read-out needed active cooling. This led to the believe that Timepix chip cameras (“Laural” and “Hardy”) should be actively cooled by Peltier water-cooling systems. Peltier cooling in vacuum is difficult to implement, therefore the Peltier cooling system was mounted outside the vacuum. This system made it possible to minimize any DAC fluctuations and to cool a system to be stable within 0.1°C at any outputs between 4W to 30W in the vacuum. (Figure 2.4 shows the test set-up and results at different power loads).

### 2.3.2 YEOL 1010 Medipix camera

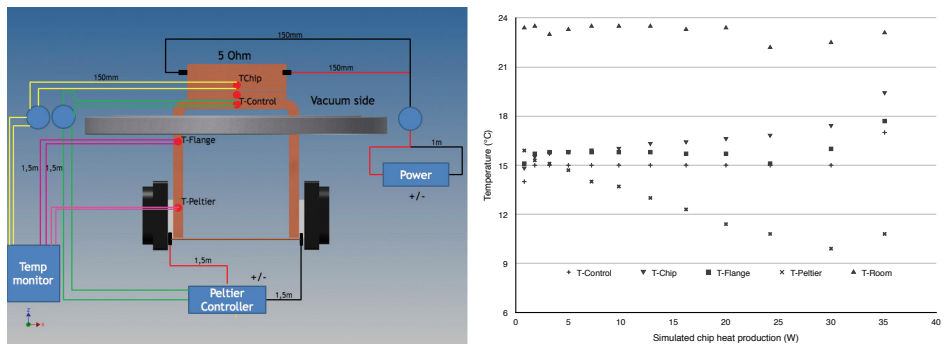
This camera is build into the image plate chamber of the YEOL 1010 by modifying the original bottom flange to support a copper-through-flange cooling block. A “snap black” supports the Nikhef Medipix2 carrier board and the hot parts of the USB2.0 FITPIX are cooled directly to the main copper pylon supporting the detector. With the active water-cooling added to the bottom of this pylon, the read-out board and detector were within design specifications at a maximum of 40° (Fig. 2.5).

### 2.3.2.1 Chip & read-out modifications

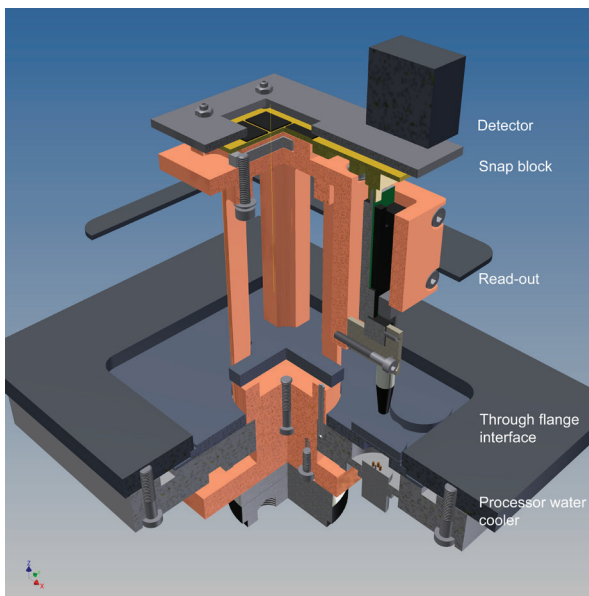
The Nikhef quad carrier board was powered through the USB read-out board. Tests showed that the power supply was insufficient for stable operation of the Quad Medipix-2 detector. This problem was solved by supplying the power directly to the carrier board by two external lab power supplies.

### 2.3.2.2 Software

Pixelman was used to acquire images and to send acquisition data to the detector (DAC and exposure times) (Turecek et al., 2011). The software was licensed with the purchase of the USB read-out board from Prague. (Kraus et al., 2011). The data was acquired on a computer that was fitted with a Intel Celeron Dual Core 2400 GHz chip and 4 GB of RAM.



**Figure 2.4: Cooling test set-up and results.** (Left) By using a 5 Ohm resistor and isolating it in vacuum, all heat needs to be removed in the same way as the heat produced in the chips. (Right) Temperature test results. While still a stable temperature is reached at 30W heat production the temperature can not be kept at the set temperature of 17 °C.



**Figure 2.5: 3D 90° cut of the YEOL 1010 camera housing.**

### 2.3.3 Quad Timepix for a Titan Krios “Hardy”

The camera was designed to be positioned at the bottom of a Titan Krios electron microscope at the “Eagle” position. Important design requirements were vibration, weight and cooling.

#### 2.3.3.1 Chip & read-out

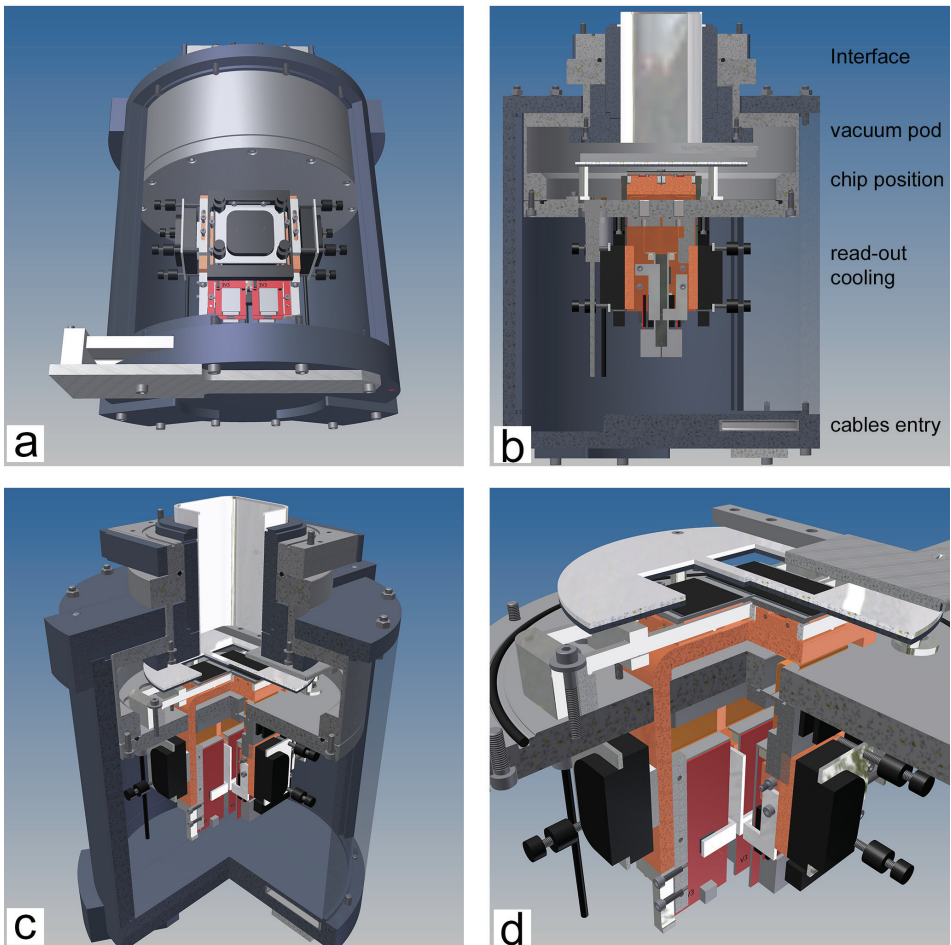
This design supports multiple Timepix Quad detectors with 300  $\mu\text{m}$  Silicon sensor layer. The chip was mounted on a Nikhef carrier board and with a Kapton flex cable connected to the Relaxd read-out board (Visser et al., 2011). The carrier board was glued to a gold coated copper snap block.

#### 2.3.3.2 Pod design

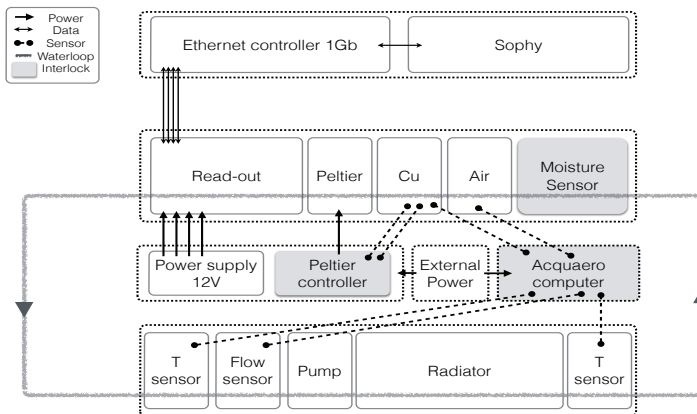
The camera pod was designed in such a way that all components are connected to the bottom flange of the pod, therefore it is possible to redesign the bottom flange to support other (new) detectors. The lead shielding and vacuum pod can then be re-used and a new flange can be designed (figure 2.8d). To support the chips of the detector and to cool them a copper bridge was vacuum soldered in the stainless steel bottom flange. The copper bridge was soldered directly to the RVS flange without using any heat isolating material (for example: ceramics). This is possible because the thermal conductivity ( $k$ ) of copper (401 W/mK) is much higher than that of stainless steel (RVS 430F) (16 W/mK). Therefore if the copper is sufficiently cooled the stainless steel will have almost no influence on the active cooling.

The lead pieces were designed according to the specifications of FEI company (FEI, 2015). A maze structure prevents any secondary X-rays to exit the lead shield. The lead shield has a minimum thickness of 15 mm in all directions and made of a special alloy (with 5% Antimony, Sb; can be tooled and is stronger than ‘pure’ lead). The bottom shield part was designed to feed all cables and water cooling hoses from the outside to the detector. A 3D printed piece was implemented in the lead shield for easy cable management.

The chips were clamped in pairs on the copper bridge and were connected with flex cables to the read-out boards. The flex cables were placed on the bottom support flange in custom small flanges. The flex cables were glued in the flanges with stycast 2850 and hardener 11. A shield of 1mm lead with a top layer of 1mm of aluminum covers everything except the detector layer. The read-out boards were cooled by two separate CPU water coolers without a Peltier element. This whole was isolated from the copper bridge which cools the Timepix chips. (Figure 2.6 shows different 3D cuts of the camera design)



**Figure 2.6:** 3D design drawings of the “Hardy” Krios camera. (a) Bottom to top view of the read-out and cooling of the non vacuum part. (b) 180 Degree cut. (c) 90 Degree cut. (d) 90 Degree cut of the bottom vacuum flange supporting all detector parts and cooling. This part can easily be redesigned to mount other detectors or newer versions of the Medipix brand detectors.



**Figure 2.7:** Design schematic for: cable management, sensors, data transport and interlocks



#### 2.3.3.3 Cooling

The water copper parts were mounted with four custom XSPC Raystorm CPU cooling blocks mounted in pairs for both the Timepix chips as well as the read-out boards. The read-out boards were directly water-cooled by the cooling blocks and the detectors were cooled by a Peltier element. The Peltier elements (Multicomp MCPF-071-14-11) had a  $\Delta T$  of 70° maximum at a voltage of 8.8V. The Peltier system with water cooling was able to cool 60W to a stable 3 degrees below ambient temperature. Tests showed that this set-up was able to keep the chips cooled within 0.1°C from the required temperature (Fig 2.5). The water is cooled to RT by four 120mm fans mounted on a Black Ice GT Xtreme 560 radiator and was pumped by a Laing DDC-1 Plus MCP355 12V Pump.

#### 2.3.3.4 Safety

Because the detector and read-out requirement is to be always powered, interlocks were implemented in case of failure. If any of the interlocks is triggered the full system will go in an emergency shutdown. There are three interlocks: (i) a moisture meter inside the lead shielding. Because moisture can form on cooled parts this value should be less than 70% when  $\Delta T$  between the air and cooled areas is ~5 °C; (ii) a computer with software to monitor water flow, temperature and fans. If a value is reaching a preset level but also if the computer and/or power fails the interlock will open; (iii) power control and temperature control of the Peltier regulator. If the power fails or the temperature at the chip position reached a preset level (30°) the interlock will be triggered. A schematic of the cable management, interlocks etc. can be found in figure 2.7.

#### 2.3.3.5 Software and Hardware

The software 'SoPhy' was supplied by ASI (ASI, 2015). It is robust and has easy access to DACs. When it is used in a Real Time Linux environment the maximum speed of the detector is 80 fps for 4 synchronized Quads or 120 fps for a single Quad. The operating system and software were installed on a 12 core 2.4GHz intel Xenon Dell 7200 computer with an SSD and 16GB of RAM. The system was connected to the read-out boards by an intel 4 port 1Gbit Intel I340-T4 ethernet card.

---

### 2.3.4 Single quad Timex design “Laurel”

This design was based on the “Hardy” design. It could be transported to different locations and to be used on different microscopes, hence simplicity. This camera supports a single Quad Timepix chip. The initial design also contains no fail saves apart from a manual temperature monitor. The acquisition software SoPhy was used and the custom build computer supports a single i7 quad core at 2.8GHz, 8 GB of RAM.

#### 2.3.4.1 Pod design

The camera is of similar design as the “Hardy” camera. The snap block of the chip was redesigned so the detector can be mounted more easily to the copper cooling bridge. Since only one chip needs to be cooled a single copper pylon was vacuum soldered in the flange. The flex cable flange is redesigned to hold only one flex cable and instead of using an isolated cooling block the read out board is cooled directly to the copper bridge. The whole system was cooled by one Peltier assembly. The lead pieces have been redesigned for simplicity. A single cone of lead was designed to fit around the vacuum chamber. The cables and water hoses then are feed through a final lead piece with a maze structure. The interface to the microscopes is easy to redesign and a redesign for a different microscope usually involves a different interface flange and lead mantle pieces (figure 2.10).

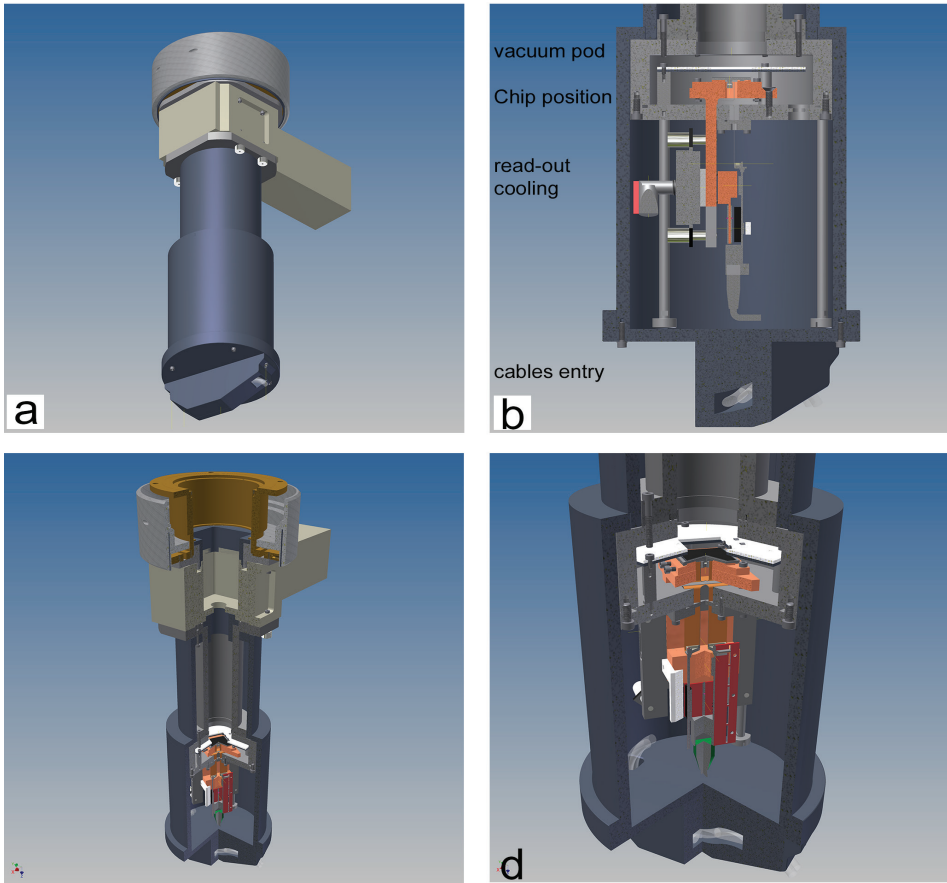
#### 2.3.4.2 Cooling

The system is cooled by a single Peltier (Multicomp MCPF-071-14-11), XSPC Raystorm CPU water-cooler, Black Ice GT Xtreme 280 radiator and a Laing DDC-1 Plus MCP355 12V Pump.

## 2.4 Discussion

The designs described in this chapter can still be improved (e.g. simplicity, access to the chip assemblies, alignment). The new Timepix-3 detector require a complete redesign of the camera housing. Especially when it comes to radiation safety, careful and precise design decisions need to be made.

In early iterations, detector failures made it clear that stable cooling and power are very important to prevent electronic failures. Cooling and power stability have a big impact on data quality. It is possible to simplify the “Hardy” design by improving the alignment of different parts (e.g. lead shielding and chip positions). Also, the read-out area could be improved by using a more simplified design. The “Laurel” design should support the same kind of safety features for emergency shut-down.

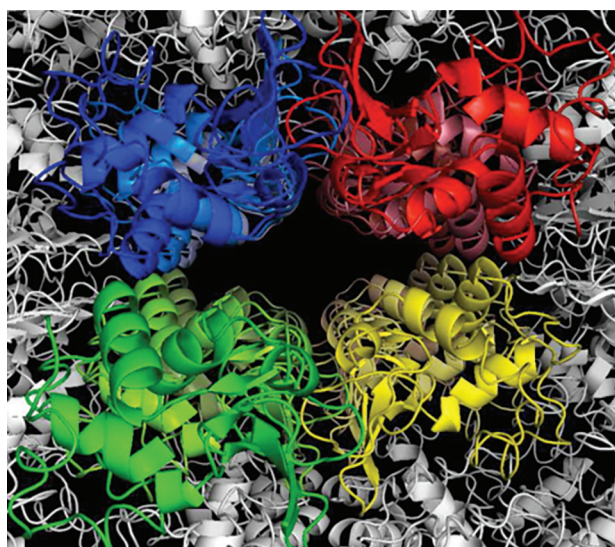


**Figure 2.8: 3D design drawings of the “Laurel” Juelich camera.** (a) Bottom to top view of the read-out and cooling of the non vacuum part. (b) 180 Degree cut. (c) 90 Degree cut. (d) 90 Degree cut of the bottom Vacuum flange supporting all detector parts and cooling.



## Chapter 3

### Protein crystal growth



Igor Nederlof, Eric van Genderen, Flip Hoedemaeker, Jan Pieter Abrahams and Dilyana Georgieva.

*InTech: Modern Aspects of Bulk Crystal and Thin Film Preparation*, ISBN 978-953-307-610-2, 01/2012, 463-476

## Contents

<b>3.1 Introduction .....</b>	<b>43</b>
3.1.1 Crystallization of bio-macromolecules.....	44
3.1.2 Purity of bio-macromolecular samples .....	45
3.1.2.1 Contaminants.....	45
3.1.2.2 Proteolysis .....	46
3.1.2.3 Sample purity tests .....	48
3.1.3 Solubility, supersaturation and phase transition .....	49
<b>3.2 Crystallization strategies.....</b>	<b>49</b>
3.2.1 Crystallization screens .....	49
3.2.2 Choosing the crystallization method .....	50
3.2.3 Vapor diffusion.....	52
3.2.4 Batch crystallization methods.....	53
3.2.5 Crystallization in gels.....	54
3.2.6 Dialysis methods .....	55
3.2.7 The role of heterogeneous substrates in the process of protein nucleation and crystallization .....	56
3.2.8 Combining heterogeneous crystallization and high throughput methods .....	57
<b>3.3 Lab automation.....</b>	<b>58</b>
3.3.1 Automation in dispensing .....	58
3.3.2 Automated experiment imaging .....	58
3.3.3 <i>In situ</i> crystal analysis .....	59
3.3.4 <i>In situ</i> diffraction analysis .....	59
3.3.5 UV detection of protein crystals .....	60
3.3.6 Second harmonic microscopy .....	60
3.3.7 Automated diffraction analysis .....	61
<b>3.4 Conclusions .....</b>	<b>61</b>

**Chapter image:** crystal structures are beautiful. One can only imagine the fluid dynamics and interactions which make such beautiful structures possible. (Fig 3.1)

---

---

## 3.1 Introduction

The biological activity of most proteins is determined by their 3D structure. For instance, a substantial number of molecular diseases are caused by protein structural alterations, which are genetically encoded. Drugs operate by binding to proteins, inducing alteration of their functional structure and thereby affecting their biological activity. Hence the design and improvement of drugs is greatly facilitated by knowledge of the 3D structures of their macromolecular targets. In the light of these considerations, it is clear that elucidation of the 3D structure of proteins is of prime importance for understanding the underlying mechanisms of molecular diseases. It was initially believed that any protein that could be made soluble and could be purified would be relatively easy to crystallize. However, the results have indicated that solubility and purity of proteins, although being important factors, do not secure a yield of useful crystals. The crystallization behavior of proteins turns out to be very complex.

In an effort to identify the naturally occurring protein folds, large structural genomics consortia were set up. The somewhat disappointing outcome of these efforts is that only about 3% of all proteins that were targeted by these consortia yielded a crystal structure (<http://targetdb.pdb.org/statistics/TargetStatistics.html>), despite massive investments in high-throughput, automated protein production, purification and crystallization. It is clear that in order to improve the current situation, better strategies for protein crystallization are required, combined with techniques that allow the use of smaller nano-crystalline material.



### 3.1.1 Crystallization of bio-macromolecules

Biocrystallization involves the three classical steps of nucleation, growth, and cessation of growth, even though the protein crystals contain on average 50% of disordered solvent (Figure 1) (Matthews, 1968). However, crystal growth of biological molecules differs substantially from small molecule crystallogenesi. The reason is the much larger number of parameters involved in biocrystallization, as well as the specific physico-chemical properties of the biological compounds. The main difference from small molecule crystal growth is the conformational flexibility and chemical versatility of macromolecules and their greater sensitivity to external factors. An overview of different parameters affecting the crystallization of biomacromolecules is presented in Table 1 (Bergfors , 2009). Another important prerequisite for successful crystallization is the quality of the macromolecular samples. Bio-macromolecules are extracted from living cells or synthesized in vitro and they are frequently difficult to prepare at a high degree of purity and homogeneity. Besides traces of impurities, the different treatments proteins are subjected to may decrease their stability and activity through different kinds of alterations. As a general rule, purity and homogeneity are regarded as conditions of prime importance. Accordingly, purification, stabilization, storage and handling of macromolecules are other essential steps prior to crystallization.

**Table 3.1: Parameters influencing the crystallization of biomacromolecules.**

<b>Intrinsic physico-chemical properties Supersaturation</b> <ul style="list-style-type: none"> <li>• Temperature, pH</li> <li>• Ionic strength and purity of chemicals</li> <li>• Pressure, electric and magnetic fields</li> <li>• Vibration and sound</li> </ul>	<b>Biochemical and biophysical parameters</b> <ul style="list-style-type: none"> <li>• Sensitivity of conformation to physical parameters</li> <li>• Binding of ligands</li> <li>• Specific additives</li> <li>• Aging of samples</li> </ul>
<b>Biological parameters</b> <ul style="list-style-type: none"> <li>• Rarity of biological macromolecules</li> <li>• Bacterial contaminants</li> <li>• Biological sources of organisms and cells</li> </ul>	<b>Purity of macromolecules</b> <ul style="list-style-type: none"> <li>• Macromolecular contaminants</li> <li>• Sequence (micro) heterogeneity</li> <li>• Conformational (micro) heterogeneity</li> <li>• Batch effects</li> </ul>

---

### 3.1.2 Purity of bio-macromolecular samples

The concept of purity has a special meaning when biological crystallogenes is concerned. Molecular samples need to be not only chemically pure, but they must also be conformationally uniform (Giegé, 1986). This concept is based on the fact that the best crystals are grown from solutions containing well-defined entities with identical physicochemical properties. For X-ray crystallographic studies, the aim is to grow 'single crystals' diffracting to high resolution with a low mosaicity and prolonged stability in the X-ray beam. It is therefore understandable that contaminants may compete for sites on the growing crystals and generate lattice errors leading to internal disorder, dislocations, poor diffraction or early cessation of growth (Vekilov and Rosenberger, 1996). Because of the high molecular weight of molecules in a single crystal (up to millions of daltons), and hence low molarity of their solutions even relatively small amounts of contaminant may induce formation of nonspecific aggregates, alter macromolecular solubility, or interfere with nucleation and crystal growth (Skouri et al., 1996; McPherson et al., 1996). Successful crystallization of rare proteins and nucleic acids support the importance of purity and homogeneity (Wierenga et al., 1987; Aoyama, 1996; Douna et al., 1993). Usually most of the contaminants are eliminated during the different purification steps, however traces of polysaccharides, lipids or proteases may still be present and hinder crystallization. Small molecules, like peptides, oligonucleotides, amino acids, as well as uncontrolled ions should also be considered as contaminants. Buffering molecules remaining from a purification step can be responsible for irreproducible crystallization. For instance, phosphate ions are relatively difficult to remove and may crystallize in the presence of divalent cations ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ). Counter ions play a critical role in the packing of biomolecules. Often macromolecules do not crystallize or yield different habits in the presence of various buffers adjusted at the same pH.

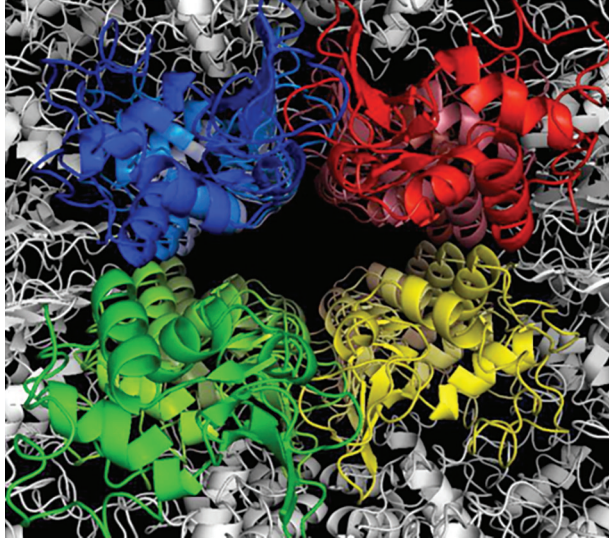
#### 3.1.2.1 Contaminants

Bio-molecular samples containing traces of contaminants can further be subjected to purification through recrystallization, column chromatography, ultracentrifugation, fractionated precipitation, affinity purification and other techniques. Microheterogeneity in pure macromolecules can only be revealed by very sensitive methods. The most common causes for heterogeneity are uncontrolled fragmentation and post-synthetic modification.

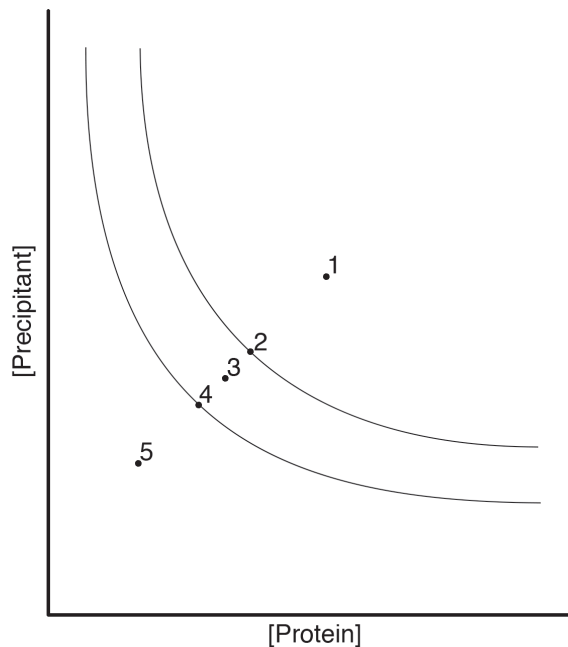
### 3.1.2.2 Proteolysis

Proteolysis normally takes place in many physiological processes and represents a major difficulty that needs to be overcome during protein extraction from the living cells that produce the desired protein (Achstetter and Wolf, 1985; Barrett and Kirschke 1981; Dalling, 1986; Bond and Butler, 1987; Arfin and Bradshaw, 1988; Wandersman 1989). The reason is that proteases are localized in various cellular compartments or excreted in the extracellular medium. Upon cell disruption, cellular compartments are mixed with extracellular proteases and control over proteolysis is lost. Decrease of protein size and stability, modification of their charge or hydrophobicity, partial or total loss of activity are usually signs of proteolysis. Traces of protease may not be detectable even when overloading electrophoresis gels, but they can cause damage during concentration or storage of samples. Poor post-translational enzymatic modifications generate micro-heterogeneity in proteins when different groups, for instance oligosaccharide chains, occupy specific modification sites on the protein, or when correct modifications are unevenly distributed. Only certain modifications are reversible, for instance phosphorylation, but others like glycosylation or methylation are not. Microheterogeneity can also appear during storage, for instance by deamidation of asparagines or glutamine residues is a well-documented phenomenon.

Pure, chemically uniform macromolecules can be fully functional in a biochemical activity assay even though they are micro-heterogeneous. Conformational heterogeneity may have several origins: binding of ligands, intrinsic flexibility of molecular backbones, oxidation of cysteine residues or partial denaturation. In the first case, macromolecules should be prepared in both forms, the one deprived of and the other saturated with ligands. In the second case, controlled fragmentation may be helpful. In the last one, oxidation of a single cysteine residue leads to complex mixtures of molecule species for which the chances of growing good crystals are low (Van der Laan et al., 1989).



**Figure 3.1: Crystal packing in lysozyme crystals (pdb:1Lyz).** This shows large cavities. These cavities are filled with disordered solvent (not shown).



**Figure 3.2: Solubility curve of a protein.** The phase state of the protein is plotted against the concentration of both protein and precipitant. At the point (1), the protein may precipitate so fast that an amorphous precipitate or at best shower of microcrystals is formed. At (2) the conditions are just right for the protein to form a stable crystal nucleus, which will start to grow – passing (3) – into a stable protein at equilibrium with the mother liquor (4). At (5), the concentration of protein and precipitant are too low for crystal nucleation or growth, and the solution will remain clear. Note that the true solubility curve of any protein is highly multidimensional, with every parameter affecting protein solubility (cf. Table 1) representing a different independent axis.

### 3.1.2.3 Sample purity tests

Although macromolecules may crystallize readily in an impure state (Holley et al., 1961), this is an exception and it is always preferable to achieve a high level of purity before starting crystallization experiments. In order to gain more information about the quality of the protein samples, different techniques can be used. For instance, spectrophotometry and fluorometry give information about the quality of samples if macromolecules or their contaminants have special absorbance or emission properties. SDS-PAGE indicates the size of protein contaminants, but not that of non-protein contaminants. Isoelectric focusing gives an estimate of the pI of protein components in a mixture and electrophoretic titration shows the mobility of individual proteins as a function of pH. The latter method can also suggest the type of chromatography suitable for further purification. Capillary electrophoresis is well adapted for purity analysis (Karger et al., 1996). Amino acid composition and sequencing of N and C termini verify in part the integrity of primary structure.

Electrospray ionization and matrix-assisted laser desorption/ionisation mass spectrometry are also powerful tools in the analysis of recombinant protein chemistry. Nuclear magnetic resonance can detect small size contaminants and gives structural information on biomolecules (Wuthrich , 1995).

It is widely believed that the success of crystal trials is largely dependent on various, not very well identified, properties of the protein. For example, a positive correlation has been established between the degree of protein monodispersity in solution and the ability of the protein to crystallize. On the other hand, it's thought to be a negative correlation between the degree of disorder in the protein and its ability to crystallize (Mikol et al., 1989, 1990).

A number of biophysical techniques and methods are employed to evaluate the quality and stability of protein solutions. Dynamic light scattering is a useful tool for non-invasive *in situ* monitoring of crystallization trials because it detects the formation of aggregates or nuclei before they become visible under a light microscope (Berne and Pecora, 1976). Fluorescence and light scattering are helpful to rapidly identify stabilizing conditions compromising simple agents (salts, co-factors etc.). Emission fluorescence is used to measure changes if the protein unfolds or undergoes other conformational changes (Konev, 1967).

---

### 3.1.3 Solubility, supersaturation and phase transition

Biological macromolecules follow the same thermodynamic rules as inorganic or organic small molecules concerning supersaturation, nucleation and crystal growth. However, protein macromolecules are organized in tertiary and quaternary structures. The intramolecular interactions responsible for their tertiary structure, the intermolecular interactions involved in the crystal contacts, and the interactions necessary to solubilise them in a solvent are similar.

To crystallize a biological macromolecule, its solution must have reached supersaturation, which is the driving force for crystal growth. The under- and supersaturated states are defined by the solubility of the macromolecules. When the concentrations of the crystallization agent and the macromolecules correspond to the solubility condition, the saturated macromolecule solution is in equilibrium with the crystallized macromolecules. Below the solubility curve (Fig. 2) the solution is under saturated and the system is thermodynamically stable. In this case, phase transition (crystallization) will not occur. Above the solubility curve, the concentration of the biological macromolecules is higher than the concentration at equilibrium. A supersaturated macromolecular solution contains an excess of macromolecules that will appear as a solid phase until the macromolecular concentration reaches the solubility value in the solution. The higher the supersaturation, the faster this solid phase appears. However, at very high supersaturation precipitation, not crystallization occurs, but insoluble macromolecules rapidly separate from the solution in an amorphous state.

## 3.2 Crystallization strategies

### 3.2.1 Crystallization screens

Finding crystallization conditions for a new protein target is largely based on a trial and error method. The first step is to set up screening trials, exposing the protein to a variety of agents in order to find useful “leads”, which can be crystals, crystalline precipitates and phase separation that point to conditions that are conducive to crystallization.

The most popular screens to perform the initial screening step are called sparse-matrix screens. These screens rely on a compilation of conditions that had previously led to successful crystallization. Systematic screens sample the crystallization parameter space in a balanced, rational way using information on the protein properties. Systematic screens are usually used as second remedy or in order to optimize the crystallization conditions.

### 3.2.2 Choosing the crystallization method

There are different methods to crystallize biological macromolecules. However, all of them aim at bringing the solution of macromolecules to a supersaturation state (McPherson, 1985; Giegé, 1987). It's important to keep in mind that not only the various chemical and physical parameters influence protein nucleation and crystallization, but also the method of crystallization. Therefore, it's wise to try different methods when searching for optimal crystallization conditions. As solubility is dependent on temperature (it could increase or decrease depending on the protein), it's highly recommended to perform crystallization trials at constant temperature unless temperature variation is part of the experiment. Solubility of most chemicals is given in Merck Index. The chemical nature of the buffer is an important parameter for protein crystal growth. It must be kept in mind that the pH of buffers is often temperature dependent, this is particularly significant for Tris buffers. Buffers, which must be used within one unit from their pK value, are well described in textbooks (Perin et al., 1974).

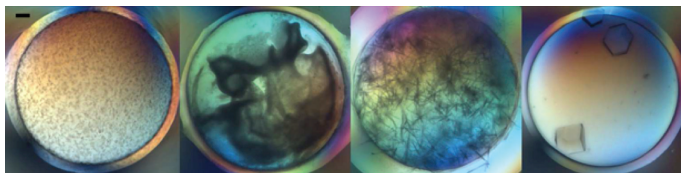
Protein samples often contain large amount of salts of unknown composition when first obtained. Thus it's wise to dialyse a new batch of a macromolecule against a large volume of well-characterized buffer of given pH, to remove unwanted salts and to adjust the pH. Starting from known conditions helps to increase the reproducibility.

Whatever the crystallization method used, it requires high concentration of biological macromolecules as compared to normal biochemistry conditions. Before starting a crystallization experiment, a concentration step is generally needed. It's also important to keep pH and ionic strength at desired value, since pH may vary when the concentration of macromolecules increases. Also, low ionic strength could lead to early precipitation. Many commercial devices are available based on Different concentration principles such as concentration under pressure, using centrifugation, or lyophilisation. The choice of method for concentration depends on the quantity and the stability of the macromolecules.

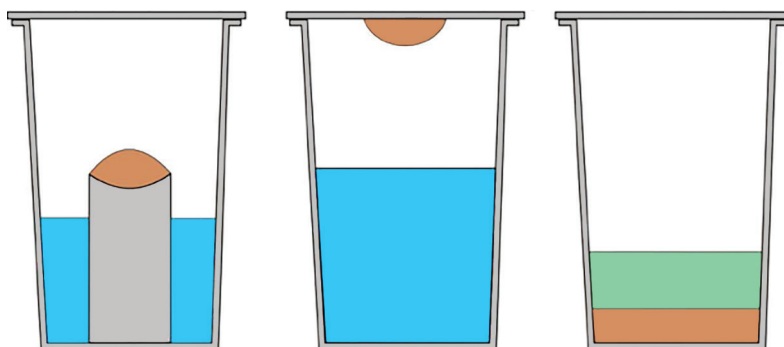
Before a crystallization experiment, solid particles such as dust, denatured proteins, and solids coming from purification columns or lyophilization should be removed. This could be achieved by centrifugation or filtration, depending on the available quantity.



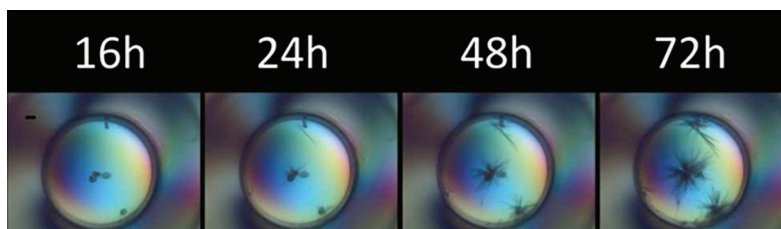
The most common method to measure macromolecular concentration is to sample an aliquot, dilute it with buffer, and measure absorbance at 280 nm for proteins within the linear range of a spectrophotometer. Proper subtraction with the reference cell should be made especially when working with additives absorbing in the 260-300 nm wavelength range. When working with enzymes, an alternative method to measure the concentration of protein is to perform activity test, otherwise colorimetric methods can be performed.



**Figure 3.3: Glucose isomerase at different crystallization conditions.** Yielding phase separation (far left) amorphous precipitation (near left) micro-crystals (near right) and macro-crystals (far right). Bar on the top left represents 200 micrometer.



**Figure 3.4: Schematic of different crystallization methods.** Sitting drop (left), hanging drop (middle), and batch crystallization (right). Well solution is blue, protein mixed with well solution is brown and oil is green.



**Figure 3.5: Lysozyme needle crystals growing on sliced human hair.** The hair acts as a nucleant, the black bar in the left picture represents 200 micrometer.



### 3.2.3 Vapor diffusion

The most widely used method of crystallization is vapor diffusion. The protein solution is a hanging, sitting or sandwich drop that equilibrates against a reservoir containing crystallizing agents at either higher or lower concentration than in the drop. Equilibration proceeds by diffusion of the volatile species (water or organic solvent) until vapor pressure in the droplet equals the one of the reservoir. If equilibration occurs by water exchange from the drop to the reservoir, it leads to a droplet volume decrease. Consequently, the concentration of all constituents in the drop will increase. For species with a vapor pressure higher than water, the exchange occurs from the reservoir to the drop. In such a 'reverse' system, the drop volume will increase and the concentration of the drop constituents will decrease. The same principle applies for hanging drops, sitting drops and sandwich drops. Most people use a ratio of 1:1 between the concentration of the crystallizing agent in the reservoir and in the droplet. This is achieved by mixing a droplet of protein at twice the desired final concentration. When no crystal or precipitate is observed, either supersaturation is not reached or one has reached the metastable region. In the latter case changing the temperature by a few degrees is generally sufficient to initiate nucleation.

Although unique in this respect, vapor diffusion permit easy variation of physical parameters during crystallization, and many successes were obtained by modifying supersaturation by temperature or pH changes. With ammonium sulphate as the crystallizing agent, it has been shown that the pH in the droplets is imposed by that of the reservoir. Consequently, varying the pH of the reservoir permits gentle adjustments of that in the droplets. From another point of view sitting drops are well suited for attempting epitaxial growth of macromolecule crystals on appropriate mineral matrices. In other words vapor diffusion provides a way to sample the crystallization space with the conditions continuously varying, as the equilibration proceeds.

---

The kinetics of water evaporation determines the kinetics of supersaturating and accordingly affects nucleation rates. Evaporation rates from hanging drops have been determined experimentally in the presence of ammonium sulphate, PEG, MPD and NaCl as crystallizing agents. The main parameters that determine the rate of water equilibration are temperature, initial drop volume, water pressure of the reservoir, and the chemical nature of the crystallization agent. Theoretical modeling has shown in addition the pivotal role of the drop to reservoir distance. It was shown that the effect of this parameter is negligible in classical set-ups and becomes only noticeable when drop to reservoir distance is more than 2 cm. From practical point of view, the time for water evaporation to reach 90% completion can vary from about 25 hours to more than 25 days. The fastest equilibration occurs in the presence of ammonium sulphate and the slowest in the presence of PEG. Equilibration rates are significantly slowed down by increasingly appropriately the distance between the drop and the reservoir. An alternative solution to decrease equilibration rates is to apply a layer of oil over the reservoir.

### 3.2.4 Batch crystallization methods

Another routinely used method for crystallization is the batch method. The biological macromolecule to be crystallized is mixed with the crystallizing agent at a concentration such that supersaturation is instantaneously reached. Crystallization trials are dispensed and incubated under low-density paraffin oil. The crystallization drops remain under oil, where they are protected from evaporation, contamination and shock. Since supersaturation is reached at the start of the experiment, nucleation tends to be higher, if compared to the vapor diffusion method. However, in some cases fairly large crystals can be obtained when working close to the metastable region. Although the microbatch method has not been compared in a statistically significant scale against hanging drop - vapor diffusion method, a comparison on a small scale has been performed (Baldock et al., 1996). The study demonstrated that the methods are not entirely identical, but are equally effective. The results suggest that vapor diffusion method and the microbatch technique will probably produce similar numbers of crystals, but may not produce crystals for the same conditions. Microbatch and vapor diffusion methods are both suitable for high throughput crystallization experiments where all the steps of dispensing, mixing and sealing are automated and performed by a robot. Other crystallization methods worth mentioning, although with more limited success and use are crystallization in gel, dialysis, microfluidics, free interface diffusion. Microfluidic chips are also being used for high throughput crystallization screening.

### 3.2.5 Crystallization in gels

Special attention has been paid to crystallization in gels (Robert and Lefauchaux, 1988). The protein crystallization process consists of two main steps – the transport of growth units towards the surface of the crystals and second, the incorporation of the growth units into a crystal surface position of high bond strength. The whole growth process is dominated by the slower of these two steps and is either transport controlled or surface controlled. The ratio between transport to surface kinetics, which can be tuned by either enhancing or reducing transport processes in solution, was shown to control the amplitude of growth rate fluctuations. These are the reasons why gels if properly designed are expected to enhance the quality of crystals. It's worth mentioning that crystals growing in gel do not sediment as they do in free solution. They develop at the nucleation site, sustained by the gel network. For small molecule crystals grown in silica gel, the gel often forms cusp-like cavities around the crystal and a thin liquid film that reduces contamination risk, separates the crystal from the gel. Such cavities have not been seen in macromolecular crystals. Recent studies have shown that silica gel can be incorporated in the crystal network almost without disturbing the crystal lattice. Such crystals that still diffract to a high resolution, are mechanically reinforced and are more resistant to dehydration, because the silica gel framework embedded in the crystal slows down water loss due to its hygroscopic properties. Although seeding can be used, it appears that most of the gel grown crystals are obtained by spontaneous nucleation inside a macroscopically homogeneous gel. When the gel adheres to the walls of the container, no nucleation occurs on the cell walls, neither on dust. So, heterogeneous nucleation is strongly reduced, if not suppressed. Another type of nucleation, namely secondary nucleation, is due to attrition of a previous crystal by the solution flux. When nucleation occurs inside the gel, one observes that all the crystals appear at the same time and have about the same size. They are homogeneously distributed in the whole volume.

---

### 3.2.6 Dialysis methods

Crystallization by dialysis methods allow for an easy variation of the different parameters that influence the crystallization of biological macromolecules. Different types of dialysis cells are used, but follow the same principle. The macromolecule is separated from a large volume of solvent by a semi-permeable membrane that gives small molecules free passage, but prevents macromolecules from circulating. The kinetics of equilibration will depend on the membrane cut-off, the ratio of the concentration, the temperature and the geometry of the cell.

The method of crystallization by interface diffusion was developed (Salemme, 1972) and used to crystallize several proteins. In the liquid-liquid diffusion method, equilibration occurs by diffusion of the crystallization agent into the biological macromolecule volume. To avoid rapid mixing, the less dense solution is poured gently on the most dense (salt in general) solution. Sometimes, the crystallizing agent is frozen and the protein layered above to avoid rapid mixing.

### **3.2.7 The role of heterogeneous substrates in the process of protein nucleation and crystallization**

In general, additives play an important role in protein crystallization. Heterogeneous substrates are usually regarded as additives when they are purposefully added to the solution in order to obtain a desired effect (inhibition of nucleation, habit change of crystals). However, impurities of foreign substances may also exist in the solution originating from other sources (the solvent, crystallization agent, etc.). Heterogeneous crystallization which is induced by a properly chosen additive may allow better control of nucleation and growth. The first report of a nucleant inducing nucleation of macromolecules was the epitaxial growth of protein crystals on minerals (McPherson and Schlichta, 1988). Other candidate nucleants followed like zeolites, silicates, charged surfaces, porous materials etc. and have been tested for multiple proteins (Sugahara et al., 2008, Takehara et al., 2008). Previous results showed that horsehair and dried seaweed showed increased hits when added to sparse-matrix crystallization trials. The increase in crystallization was 35% when horsehair was added to 10 test proteins (Thakur et al., 2008). The underlying mechanism is explained with epitaxial nucleation in the case of minerals, electrostatic interactions if the nucleants contain charged surfaces, nucleation through specific favorable protein-protein interactions or physical entrapment in the caves of porous materials.

Seeding techniques can be advantageous in both screening of crystallization conditions to obtain crystals as well in the later optimization steps. The streak seeding technique may provide a fast and effective way to facilitate the optimization of growth conditions without the uncertainty that is intrinsic in the process of spontaneous nucleation (Bergfors, 2003). A probe for analytical seeding is easily made with an animal whisker mounted with wax to the end of a pipette tip. The end of the fiber is then used to touch an existing crystal and dislodge seeds from it. Gentle friction against the crystal is normally sufficient. The probe is then used to introduce seeds into pre-equilibrated drop by rapidly running the fiber in a straight line across the middle of the drop containing protein and precipitant. Sitting drop set-ups are preferable since hanging drops tend to evaporate more quickly.

---

### 3.2.8 Combining heterogeneous crystallization and high throughput methods

A method for the introduction of heterogeneous nucleants in high throughput crystallization experiments has recently been developed (Nederlof et al., 2011). The method includes preparing of crystallization plates that are locally coated with fragments of human hair, allowing automated, high throughput crystallization trials in a fashion entirely compatible with standard vapor diffusion crystallization techniques. The effect of the nucleants was assessed on the crystallization of 11 different proteins in more than 4000 trials. Additional crystallization conditions were found for 10 out of 11 proteins when using the standard JCSG+ screen. In total, 34 additional conditions could be identified. The increase in crystallization conditions ranged between 33.3% (two additional conditions were identified for myoglobin on top of four homogeneous crystallizations) to 1.2% (we identified a single additional condition for insulin, which crystallized in 85 out of 96 conditions); the median increase in crystallization hits was 14%. The method is straightforward, inexpensive and uses materials available in every crystallization lab.

## 3.3 Lab automation

In recent years, setting up protein crystallization trials and analysis of the results has become largely automated. More and more of the crystallization methods mentioned in section 5 have been made amenable to automation, with the sitting drop method still the most popular experiment type in this respect. Lab automation includes the use of dispensing robots, imaging robots, *in situ* crystal analysis as well as automated diffraction analysis (Stevens, 2000, Berry et al., 2006).

### 3.3.1 Automation in dispensing

Dispensing robots that are used routinely are either specialized for dispensing well recipes (e.g. Formulatrix, MatrixMaker) or drop-setting (NT8, Phoenix, Mosquito), but there are also more generic robots that can do both (Hamilton Star, Tecan Evo). In general, the experimenter will start crystallization trials with a set of pre-defined conditions, contained in one or more screens. Over 150 of these screens can be bought from commercial vendors in a wide range of formats. A number of these are designed on the basis of statistical analysis of results obtained at structural genomics initiatives. When initial hits are found with screens like these, secondary optimization experiments need to be performed to produce diffracting crystals. In this stage, interaction with a Lab Information Management System (LIMS), where experiment design can be coupled to experiment preparation and analysis, greatly enhances the potential throughput in a lab and thereby the success rate. There are a number of these software packages that can be used to create grid experiments around an initial hit condition, as well as randomized sparse-matrix screens based on initial successes.

### 3.3.2 Automated experiment imaging

Automated experiment analysis is an essential part of the lab setup. Due to the increase in throughput obtained by using dispensing robots it is impossible to routinely scan the results manually under a microscope. The dynamic nature of these experiments can cause the crystallographer to miss events, even crystals. Imaging robots vary from semi-automated microscopes with a moving plate stage and camera to fully automated incubators that are capable of following all lab experiments from start to finish without human intervention. Ideally, images are displayed to the user in the context of the experiment design, so that the results are easily interpreted. If this functionality is integrated with the experiment design and dispensing the optimization circle is complete. Such LIMS systems (Bard et al., 2004) can be further expanded to follow up on harvested crystals, to assess their diffraction quality and finally the structural data derived. (see 7.4)

---

### 3.3.3 *In situ* crystal analysis

When crystals are found an assessment needs to be made whether the crystals are indeed protein crystals or just salt crystals. And the quality of the crystal needs to be established as well as their usefulness for collecting diffraction data. It has always been difficult to distinguish protein crystals from salt crystals without actually collecting diffraction data. Historically, destructive methods have been used like the “crunch” method and protein dyes, the idea being that crystals similar to the ones destroyed will have the same properties. These methods have not always been conclusive and often the true nature of the crystal was only revealed on the X-ray beam. In recent years, three new techniques have been developed in this field; *in situ* diffraction analysis, UV detection and second harmonic microscopy.

### 3.3.4 *In situ* diffraction analysis

A number of years ago, Oxford Diffraction has come up with a device for X-ray diffraction analysis of crystals in the plate where they were grown (Skarzynski 2009, le Maire, et al., 2011). The idea is fairly simple, you center a crystal in the X-ray beam using a visual alignment tool and you subsequently take a single or a small number of X-ray diffraction images to assess whether a crystal is indeed protein, and to get some idea about the diffraction quality (mosaicity, resolution). An advantage is that the method is non-invasive (bearing in mind potential radiation damage) and fast. The method is not suitable for complete diffraction analysis, as the sample can only be rotated by 6°. It is also possible to automatically screen a complete crystallization plate for potential diffraction. When suitably diffracting crystals are found they will still need to be harvested and frozen for complete diffraction analysis.



### 3.3.5 UV detection of protein crystals

An increasing number of imaging devices (see 7.2) make use of a secondary light path in the UV range to detect protein crystals. These imagers make use of the fluorescence in UV by proteins, mostly caused by tryptophan (Judge et al., 2005). Since the protein concentration in the crystal will be much larger than in solution, any protein crystal will light up under UV, provided that the protein contains tryptophan. This is a relatively fast and non-invasive method, UV illumination can cause some ionisation in the drop, but this effect is much less than with X-ray illumination. In order to maximize its use, the experiment media (plates, seals) have to be chosen with care, some plastics are not sufficiently translucent in UV, or fluoresce themselves, adding noise to the image. One also has to bear in mind that some non-protein crystals (ATP, other co-factors added), might also fluoresce in UV. Having visible light and UV cameras integrated in a single imaging device greatly enhances its usefulness to distinguish protein from salt crystals.

### 3.3.6 Second harmonic microscopy

Fairly recently, a new development in the field of *in situ* crystal analysis has been reported. The technology makes use of a phenomenon called second harmonic generation (SHG), more often referred to as “frequency doubling” (Wampler et al., 2008). When an intense laser pulse travels through a highly polarizing, non-centrosymmetric material, light emerges with exactly half of the wavelength of the incident beam. The explanation is that two photons of the incident beam merge, creating a single photon with twice the energy. If the incident beam is in the near-infrared, the emerging beam, will be in the visible range. As mentioned, the technology requires intense laser light, delivered in femtosecond pulses. Most chiral crystal classes, with the exception of octahedral and icosahedral crystals, allow for SHG, thus encompassing over 99% of all protein crystals grown so far. A first commercial device using this technology, called SONICC, is available since 2011. When combined with a LIMS and a visible light imaging station, SHG can be used to automatically score and pre-sort the results for the experimenter.

---

### 3.3.7 Automated diffraction analysis

In parallel with automation taking hold of many crystallization labs, the last part of the protein structure analysis pipeline, diffraction analysis in the X-ray beam, is increasingly automated as well. Not so long ago, a crystallographer would either measure his/her crystals at a home X-ray source or would travel to a synchrotron facility to do so. The process involved manually harvesting of the crystals, preparing them for the X-ray beam (mounting in a capillary or in a cryoloop, freezing), mounting them manually in the X-ray beam, gathering a few trial images to determine optimal settings for exposure, distance etc. and finally recording a set of diffraction images to solve the structure. The most time consuming steps have now been automated (Cork et al., 2006, Song et al 2007). Most notably, crystal mounting robots will now automatically take samples out of a liquid nitrogen dewar and place them in the X-ray beam, eliminating the need for the user to enter the X-ray hub of the synchrotron after every crystal. At the home lab, the mounted crystals are packed in specific dewar compatible with the robot arms at the beamline, and they are mailed to the synchrotron. In many synchrotrons, the user now has a choice of having a local operator collecting the data, or to drive the computers at the facility remotely from their own lab, there is no need to travel to a remote synchrotron anymore. In the near future, the automation can be improved by automatic crystal centering routines (Vernede et al., 2006). With e.g. the use of SHG (see 7.3) crystals can automatically be located inside the cryo loops, and this information can be used to automatically center the crystal in the beam.

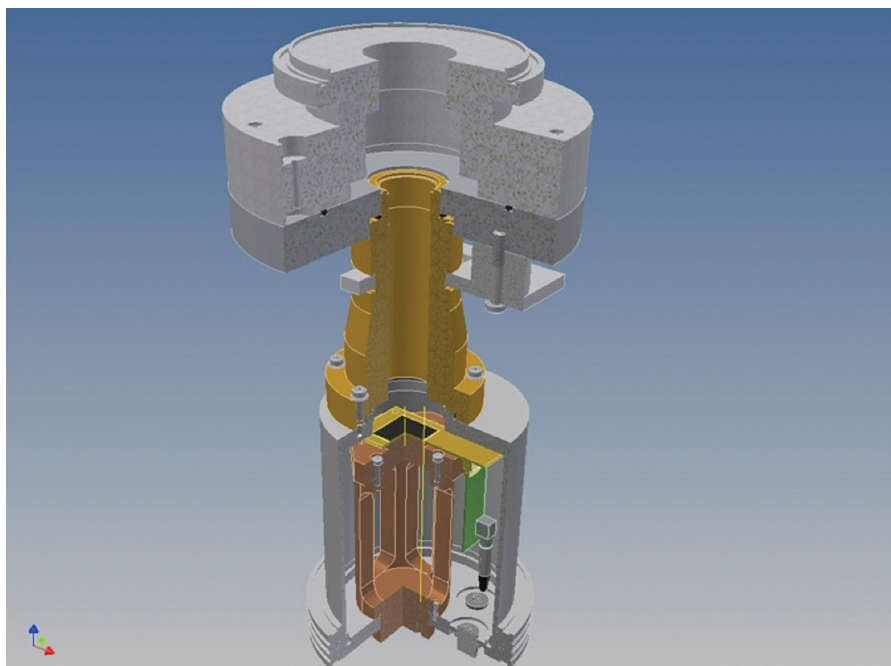
## 3.4 Conclusions

The chapter covers some of the main aspects of protein nucleation and crystallization. Different diagnostic tools, crystallization techniques and strategies are explained. New tendencies in the field such as combining heterogeneous nucleants and high throughput methods are also presented.



## Chapter 4

### **A Medipix quantum area detector allows rotation electron diffraction data collection from submicrometre three- dimensional protein crystals**



Igor Nederlof, Eric van Genderen, Yao-Wang Li and Jan Pieter Abrahams  
*Acta Crystallogr D Biol Crystallogr.* 2013 Jul; 69(Pt 7) : 1223-30

## Contents

<b>4.1 Introduction .....</b>	<b>65</b>
<b>4.2 Materials and methods .....</b>	<b>68</b>
4.2.1 Crystallization .....	68
4.2.2 Vitrification.....	68
4.2.3 Rotation electron diffraction.....	68
4.2.4 Measuring electrons using a Medipix-2 detector.....	69
4.2.5 Preparing diffraction patterns for data processing.....	72
4.2.5.1 Removing the cross .....	72
4.2.5.2 Correcting bad pixels .....	72
4.2.5.3 Centring the images.....	73
<b>4.3 Results .....</b>	<b>74</b>
4.3.1 Collecting diffraction data .....	74
4.3.2 Processing the diffraction data with MOSFLM .....	75
4.3.2.1 Beam centre.....	75
4.3.2.2 Angle of the rotation axis .....	75
4.3.2.3 Unit cell parameters .....	75
4.3.2.4 Determining the orientation of the crystal .....	76
4.3.2.5 Beam divergence .....	76
<b>4.4 Discussion .....</b>	<b>77</b>
<b>4.5 Conclusion.....</b>	<b>79</b>

**Chapter image:** The first camera which showed the potential of these cameras. The strong points of the detector and the design flaws made us to pursue even better camera designs. (Fig. 4.2)

---

---

## 4.1 Introduction

Protein crystallography is a major justification for large-scale X-ray facilities like synchrotrons and free electron lasers. However, 3D protein crystals that are smaller than about 0.5  $\mu\text{m}$  are too small for standard X-ray crystallography, although XFEL sources are expanding the method towards smaller crystals (Chapman et al., 2011). This is a serious bottleneck, as about 30% of proteins that crystallize, do not grow crystals of a sufficient size or quality for X-ray structure determination (Rupp, 2004). Especially membrane proteins and large (dynamic) protein-nucleic acid complexes fail to grow into crystals of sufficient size. Structural information on these important drug targets is therefore severely lacking.

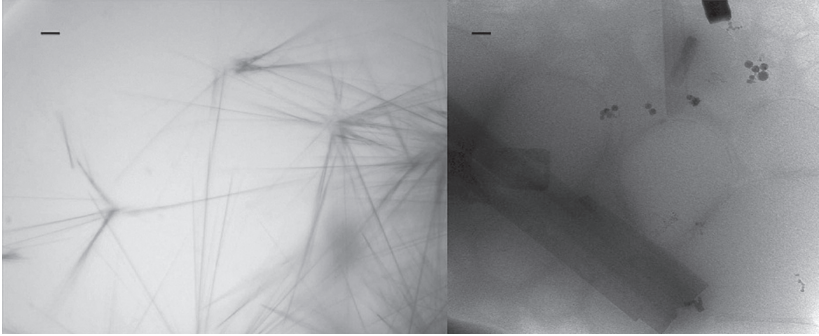
Electrons are less damaging to proteins than X-rays by several orders of magnitude per diffracted quantum (Henderson, 1995). This property of electrons explains the successes of 2D electron crystallography. For instance, 45Å thick, sub-micrometer patches of bacteriorhodopsin 2D crystals yielded images with a resolution of 2.8Å (Baldwin et al., 1988). 3D crystals with an equivalent volume would measure approximately 150x150x150 nm<sup>3</sup>. Recent data demonstrates that useful, high-resolution electron diffraction data (up to 2.5Å) can be obtained from nano-sized 3D protein crystals, where synchrotron X-rays fail (Jiang et al., 2009).

However, only single diffraction shots could be collected. Collection of rotation data of protein nano-crystals was not possible because the signal-to-noise ratio and dynamic range of CCD detectors and image plates was insufficient. Very recently, direct electron detectors became available, which have a better signal-to-noise ratio and which may be better suited. These new detectors are very expensive and are probably not sufficiently radiation-hard to be routinely exposed to the direct electron beam.

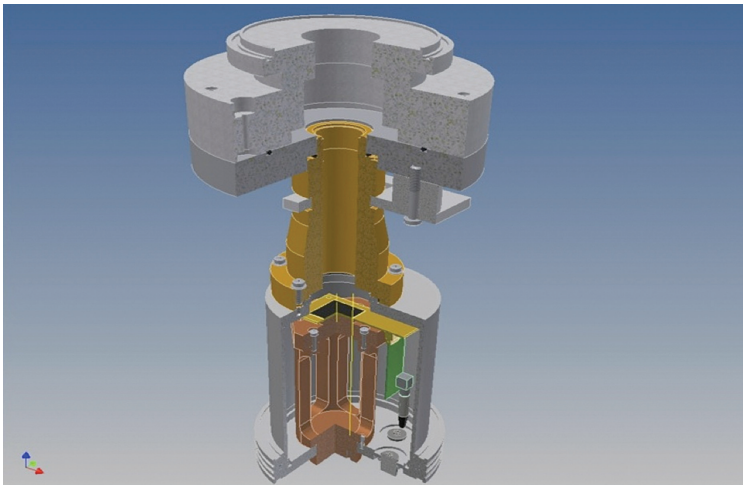
Electrons can also be detected by quantum area detectors like Medipix-2 (Llopart et al., 2002, Faruqi and Henderson., 2007). The Medipix-2 detector is more radiation hard than other direct electron detectors, like the Falcon, because its read-out electronic circuitry (which is sensitive to radiation damage and interference) is shielded by a semiconductor sensor layer, to which it is bump-bonded (6). The detector has a very high signal-to-noise ratio because each pixel has its own readout electronics that measures the hole-charges that are produced by an incident electron hitting the sensor layer within 10<sup>-5</sup> s. If the integrated energy after amplification is above a set threshold corresponding to the energy of a 200 keV electron, the incident quantum is counted as a “hit”. Thus the Medipix-2 chip only counts 200 keV electrons, and, unlike many other detectors, is blind to soft X-rays of lower energy that are also produced in great abundance inside any EM. In this fashion its noise is almost exclusively determined by the counting statistics of the electrons. This gives a significant improvement over conventional CCD cameras in electron microscopes (Faruqi and McMullan, 2011, Georgieva et al., 2011).

We recently showed that a Medipix-2 detector with a 500 nm sensor layer can detect 200 keV electrons with a signal-to-noise ratio that is at least an order of magnitude better than that of image plate (Georgieva et al., 2011). The Medipix detector is highly sensitive at low count rates, allowing accurate measurement of the high-resolution terms. In addition, with a 500 nm sensor layer it is sufficiently radiation hard to routinely be exposed to a direct beam of 200 keV electrons. It also has a high dynamic range allowing accurate measurement of the intense, highly peaked dose it receives in the low-resolution diffraction spots. For 200 keV, the point spread of the detector is increased, but if the point spread is not much higher than the spread of the Bragg spots this would not matter for collecting diffraction data.

In view of these desirable characteristics of the Medipix-2 detector, we investigated whether it would be possible to reduce the electron dose per diffraction frame by an order of magnitude and still measure high-resolution diffraction data, as this would, for the first time, allow collecting rotation diffraction data of single protein crystals.



**Figure 4.1: Lysozyme nano-crystals.** In a crystallization drop (left) and an electron micrograph of vitrified crystals (right). Scale bars correspond to 10  $\mu\text{m}$  (left) and 500 nm (right).



**Figure 4.2: Cut-out view of the Medipix2 camera assembly.** Electrons enter the tube from the top and hit the detector (black square) on the NIKHEF assembly board (Belo et al., 2003), which is connected through the USB read-out electronics and a vacuum connector to the computer.



## 4.2 Materials and methods

### 4.2.1 Crystallization

Crystallization experiments were carried out using the sitting drop vapor diffusion technique in Innovadyne SD-2 plates. Rockmaker software (Formulatrix) was used for designing the experiments. A Genesis (Tecan) robot was used for dispensing the screening solutions in the reservoirs of MRC2 plates (Swissci). The ORYX 6 (Douglas Instruments) crystallization robot was used for transferring 500 nl reservoir solution and 500 nl protein solution in sitting drop wells. Plates were stored at 20 °C and imaged using the automated imaging system Rock Imager (Formulatrix). Lysozyme (8 mg/ml) formed needle-shaped crystals after 48h when mixed 1:1 with well solution containing 0.1 M Sodium Acetate pH 3.8 and 1.0 M Potassium Nitrate.

### 4.2.2 Vitrification

Protein crystals were vitrified using a Vitrobot (FEI, 2015). 3  $\mu$ l of well solution was mixed with the drop containing nano-crystals and transferred to a 3mm glow-discharged holey carbon grid (AGAR). Excess liquid was blotted away (blot time 3 sec blot force 5) and the sample was plunge frozen in liquid ethane cooled by liquid nitrogen.

### 4.2.3 Rotation electron diffraction

Diffraction data were collected on a CM200FEG (Philips) transmission electron microscope at the National Center for High Resolution Electron Microscopy in Delft. The energy of the electrons was 200 keV (wavelength 0.025Å). Samples were cooled to -180° Celsius in liquid nitrogen in a home-modified cryo-transfer holder (Gatan, 2015). Diffraction patterns were collected using a Medipix-2 detector. CBED (convergent beam electron diffraction) was used to collect the data. The beam was hardly convergent, as indicated by the data analysis described below ( $< 0.4^\circ$ ). A small (10  $\mu$ m) condenser aperture and spot-size 11 were used to limit beam divergence and intensity, which also served to reduce crystal damage. No beam blanker was used. An automated compu-stage allowed crystal alignment and rotation. We selected crystals with a thickness in the order of 100 nm (10 to 30 unit cells). Thicker crystals had a reduced signal due to electron absorption; thinner crystals had a lower diffraction yield due to their limited diffracting volume.

---

#### 4.2.4 Measuring electrons using a Medipix-2 detector

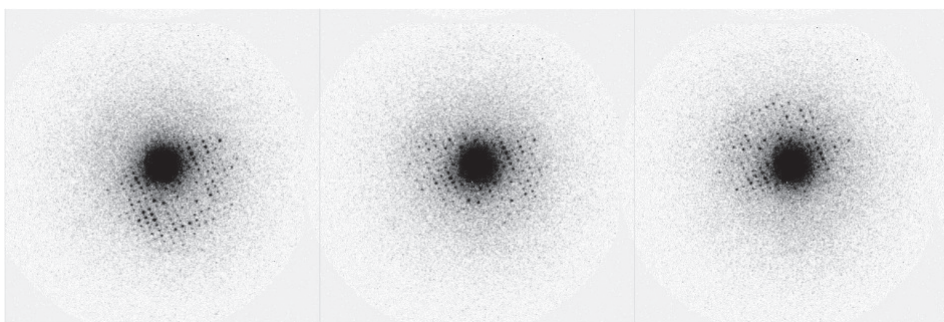
Diffraction patterns were collected with a CMOS Medipix-2 detector mounted on a Nikhef carrier (Bello et al., 2003) board and the data were transferred to a Windows PC using USB 1.1 read out electronics (Vykydal et al., 2006). Four Medipix-2 ASICs, each with 256x256 pixels and a pixel size of 55  $\mu\text{m}$  are assembled to make a Medipix-2 Quad. The Quad was covered by a single custom-made 500  $\mu\text{m}$  semiconductor sensor chip. The distance between two neighboring single Medipix-2 ASICs is approximately 250  $\mu\text{m}$ , so the edge pixels are about 125  $\mu\text{m}$  wide. Therefore they will capture more signal than the other pixels resulting in a bright cross that quarters the raw images. This cross can be corrected for as discussed below.

The detector has a big dynamic range where each pixel consists of a separate 14 bit pseudo-random counter. The test circuitry and four bit trimming system are able to compensate for most fabrication variations and therefore the overall global threshold has a variation of 95 eV. The electronics noise can account for another 110 eV. These two values combined result to a dynamic range of no lower than 900 eV between channels. In silicon it requires  $\sim 3.6$  eV to produce one electron-hole pair, therefore a 200keV electron can produce  $\sim 55,000$  pairs. This means that the dynamic range of the electronic noise accounts for less than 1.6% of the total deposited energy per electron incident and even less when there is more than one electron incident per clock cycle (McMullan et al., 2007, Plackett et al., 2010)

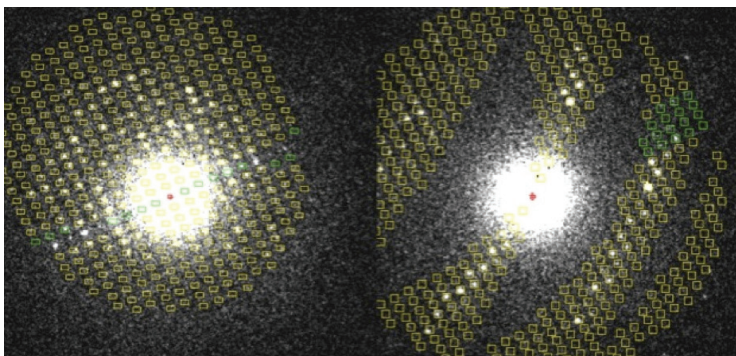
The thickness of the sensor layer (500  $\mu\text{m}$ ) is larger than for stock Medipix-2 chips (which is 300 $\mu\text{m}$ ). We selected this larger thickness to prevent the 200 keV electrons from penetrating through the sensor layer and damaging the underlying electronics. Monte Carlo simulations (300  $\mu\text{m}$  Si sensor layer; 120 keV, 200 keV and 300 keV) predicting such events have been previously described. (Faruqi and Henderson, 2007, McMullan et al., 2007, McMullan et al., 2014). Most importantly it also describes the events where these higher energy electrons scatter in the Silicon layer to neighboring pixels. As long as this scattering stays close to the boundaries of the Bragg spots, this would give no major point spread problems.

As previously described (Faruqi and McMullan, 2011) there is a need to set the threshold settings such that an electron of 200 keV is counted even when its energies are spread over multiple pixels. This is in a situation where a 120 keV electron would not be counted because the threshold is set close to the equivalent level of a 120 keV incident. As described above the chance of a single high energy electron scattering in the sensor layer to neighboring pixels increases with higher energies. To get a higher electron counting gain a compromise threshold level was chosen where also lower energy particles would be counted in each pixel. This, however is not much of influence for doing diffraction studies with the Medipix detector, since most other particles (X-ray) will stay well below an equivalent 100 keV threshold setting. Because the detector is used for diffraction studies the resulting point spread does not cause any problems since the spread will follow a Brag spot profile.

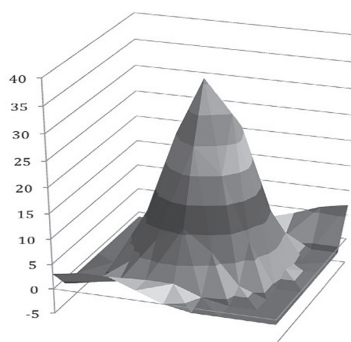
The vacuum pod in which the Medipix-2 was mounted on the on-axis port of the CM-200 FEG electron microscope is shown in fig 4.2. It houses both the carrier and USB1.1 readout board. Damage to the electronics of these boards by the direct electron beam was prevented by the small window of the pod, which only allowed illumination of the about 80% of the total area of the Medipix-2. Out-gassing of the electronics at  $1.10^{-6}$  Pa vacuum did not prove to be a problem for both the electronics and the TEM, despite 2 years of usage. The bias voltage on the Medipix-2 is set to 100V to be able to count the holes after the electron incidents (Georgieva et al., 2011). The 4 chips are set to non-parallel read-out by the Pixelman v2.0 software (Turecek et al., 2011) on a PC with windows 7 in Java mode. The Medipix-2 Quad was calibrated using standard tools which are included in Pixelman: (i) DACs were scanned to determine the noise edge of the different ASICs; (ii) threshold equalization was performed to remove small standard discrepancies between individual pixels (3 bits). This procedure adjusts each pixel so that its threshold is as close as possible to the average and was performed with the standard settings supplied with the Pixelman package; (iii) each lower threshold (THL) was set to a value close to the noise edge in order to acquire as much signal as possible from the incident electrons; (iv) flat field images obtained by illumination with a uniform beam allowed equalizing the four chips. After the initial calibration, a very low intensity uniform beam was used in combination with the integral method to further fine-tune the different chips.



**Figure 4.3: Three diffraction patterns.** Selection of a series of 23 subsequent frames ( $0.066^\circ$  per frame,  $0.083^\circ$  between adjacent images) collected at 200 keV with a Medipix-2 Quad detector (512 by 512 pixels). The crystal was orthorhombic lysozyme, maximum resolution  $\pm 1.8\text{\AA}$ .



**Figure 4.4: Overlay of predicted spot positions.** (resolution threshold  $2\text{\AA}$ ) Combined with observed diffraction patterns of lysozyme nano-crystals, indicating that the 'a'-axis could be close to  $125\text{\AA}$ , while the 'b' and 'c' axes are about  $34.2\text{\AA}$  and  $25.5\text{\AA}$ . For the Images are screen shots of MOSFLM runs.



**Figure 4.5: Spot profile from MOSFLM for the first data set.** The second data set had a similar spot profile (not shown).

## 4.2.5 Preparing diffraction patterns for data processing

The Medipix-2 images need some pre-processing before they can be read into MOSFLM. Three problems needed to be addressed: (i) the cross needs to be removed; (ii) dead pixels needed to be corrected; (iii) the image needed to be centered, as there was a substantial drift of the direct beam. To this aim, we wrote a program that solved these issues as described below.

### 4.2.5.1 Removing the cross

A bright cross appears in the raw Medipix-2 Quad images, as it is a tiled assembly of four single Medipix-2 ASICs. The edge pixels of each single Medipix-2 ASIC that abuts another in the Quad assembly are larger than the other pixels (they are 125 mm wide, rather than 55 mm). Therefore, the pixels at the interface between single Medipix-2 ASICs capture more electrons. This results in a spatial distortion and a non-uniformity of the measured signal. In order to correct for the spatial distortion, our program shifts pixels with an x-value smaller than 256 by one pixel to the left and it shifts pixels with a value of 256 or higher by one pixel to the right. It applies vertical shifts in a similar fashion. This procedure duplicates the pixels adjacent to the horizontal and vertical bisecting lines (and quadruples the four centre pixels). Their value was divided by 2.3 (or 5.3 for the quadrupled centre pixels) to correct the increased measured intensities of these pixels.

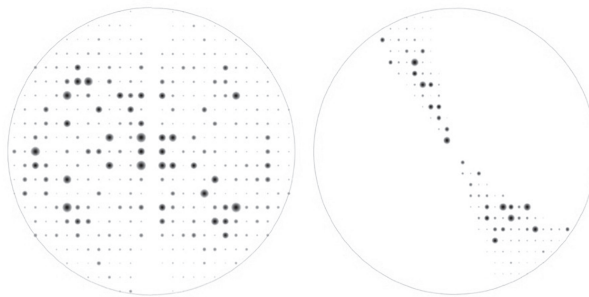
### 4.2.5.2 Correcting bad pixels

Despite tuning the pixels of the Medipix-2 detector, it can still have dead pixels (or bright pixels). Since positions of dead pixels are fixed and their values do not follow a Poisson distribution over time, it is possible to identify these dead pixels. For each pixel position, we calculated its variance over a range of images, and compare this variance to the variances at the other positions. ‘Good’ pixels will have a much higher variance than ‘bad’ pixels since the values of the latter tend to be relatively constant and largely independent of the signal. If the variance value was below a set threshold the pixel was labeled as ‘bad’. After having identified the bad pixels, every image was corrected. For every bad pixel in every image, its value will be replaced by the mean value of the good pixels that surrounded it. If a bad pixel had no good neighbors, the procedure was iterated after reclassification of the pixels.

---

#### 4.2.5.3 Centring the images

The position of the beam centre shifted as a function of the tilt of the sample holder. The shift was not huge (15 pixels maximally), but sufficiently large to interfere with data processing. Images were centered by calculating their cross-correlation functions with a 2D Gaussian function peaking at pixel (256, 256) and applying the calculated shift. After centering the images, they were written to disk in the CCP4 '.pck' format (Abrahams, 1993) as 1200 by 1200 pixel images, so that they could be processed as small MAR images by MOSFLM.



**Figure 4.6: Laue zones at  $h=0$  of the scaled data.** (as displayed by 'viewhkl'.) Note that the second data set (right pattern) does not have a complete ' $h=0$ ' Laue zone.

## 4.3 Results

### 4.3.1 Collecting diffraction data

Prior to installing the Medipix-2 detector in the EM, we collected diffraction data on image plate, film and CCD, but with these detectors we could never collect multiple diffraction patterns of single protein crystals to high resolution because of beam damage (Jiang et al., 2009). The Medipix-2 detector has an efficiency at low count rates that is at least an order of magnitude higher than that of image plate (Faruqi and McMullan, 2011, Georgieva et al., 2011). This meant that we could for the first time collect rotation data over multiple images without severe beam damage.

Unfortunately, the hardware prevented us from interfacing the detector with the electron illumination and the rotation of the sample holder. This meant in practice that we had to start rotating the crystal slowly, switch on the beam and start collecting images until the crystal had sustained too much damage. Since transferring a single Medipix-2 frame into the computer takes 0.7 s (we were using a USB1 interface), this meant that we could only collect rotation data with angular gaps between adjacent images. By using fine phi slicing, we reduced the systematic errors that were introduced by this unfortunate, but in the circumstances inevitable experimental flaw.

We collected rotation data of many crystals, improving data collection strategies and sample handling iteratively. Here we describe representative results using two rotation data sets from the same single crystal, but collected at different positions, using a fresh part of the crystal. We rotated in the positive direction for the first wedge of data and in the negative direction for the second wedge. Both wedges started at the same goniometer setting, but processing revealed their orientations to be about  $1.5^\circ$  apart. The first wedge was measured with a rotation speed of  $0.083^\circ$  /image ( $0.05^\circ$  /s), reading out an image every second. This resulted in a data set in which each image had recorded about  $0.050^\circ$  of rotation data, with a gap of  $0.033^\circ$  until the next image. For the second data set, the crystal was rotated at a speed of  $0.26^\circ$ /s, reading out an image every second. This resulted in a data set with  $0.15^\circ$  per image with gaps between images of  $0.11^\circ$ . Both data sets had diffraction spots to a resolution of  $1.8\text{\AA}$  in the early images (see figure 4.4).



---

## 4.3.2 Processing the diffraction data with MOSFLM

Several essential parameters had to be extracted from the diffraction patterns: the angle of the rotation axis relative to the detector coordinate frame, the beam centre, the beam divergence, the unit cell parameters and the orientation of the crystal. The determination of each of these parameters will be discussed in more detail below.

### 4.3.2.1 Beam centre

We observed the beam centre to shift considerably (about 15 pixels) between the starting and ending rotation angles. Since we did not use a backstop, this shift could readily be corrected for after data collection by shifting back the images accordingly (as described in the Methods section). At the time of data collection, it was unclear what the cause of the beam shift was, but later it transpired that a screw of the EM stage close to the crystals had become slightly magnetised.

### 4.3.2.2 Angle of the rotation axis

As the electrons spiral their way through the magnetic lenses of the microscope, the rotation axis that is observed on the detector is not necessarily an orthogonal projection of the physical rotation axis. It needs to be calibrated for every camera distance. Careful analysis of the diffraction patterns using the human eye, a ruler and a protractor indicated that at a virtual camera length of 700 mm and an electron energy of 200 keV, the angle between the rotation axis and the x-direction of the detector was around  $115^\circ$ . We could refine this initial estimate to  $116.5^\circ$  by overlaying predicted diffraction patterns on observed diffraction patterns once we had good estimates of the other parameters. When rotating in the negative direction, the rotation axis was redefined to be  $-63.5^\circ$ .

### 4.3.2.3 Unit cell parameters

Using our knowledge of the virtual crystal to film distance and the wavelength (which is  $0.0251\text{\AA}$  for electrons at 200 keV energy), we interactively estimated the two spacings in the plane of the detector using the ‘measure’ facility in MOSFLM. These parameters ( $34.2$  and  $25.5\text{\AA}$  with an angle of  $90^\circ$ ) were consistently found in both rotation ranges. No low index spacings of the third axis were present in the images, and its magnitude could only be estimated from the location of the lunes in the second data set. A reasonable, but by no means perfect overlay between observed and predicted spot positions could be obtained for a unit cell of  $125 \times 34.2 \times 25.5\text{\AA}$ , with all cell angles  $90^\circ$  (figure 4.5). In the past we found orthorhombic nano-crystals of lysozyme to have  $P\ 2_1\ 2_1\ 2_1$  symmetry, with a unit cell of around  $31.5 \times 52.5 \times 89\text{\AA}$  (Jiang et al., 2009). However, we could not index the rotation data with the latter cell parameters.

---



#### 4.3.2.4 Determining the orientation of the crystal

Due to radiation damage, we could only measure small wedges of data. Auto-indexing routines could therefore not cope with the data. We indexed interactively and let MOSFLM refine the orientation using post-refinement, which, despite the angular gaps between images, managed to improve the fit between observed and predicted patterns. The wedges contained sufficient data for orientation refinement, but did not allow for the refinement of cell parameters.

#### 4.3.2.5 Beam divergence

Once we had satisfactory predictions of the diffraction patterns, we adjusted the beam divergence (or mosaic spread: the two phenomena cannot be disentangled easily) in MOSFLM so that all spots were covered by the predicted pattern. The combined beam divergence/mosaic spread was about  $0.3^\circ$  to  $0.4^\circ$ .

Typical profiles of the spots are shown in figure 4.3. There were a number of bad spots, mainly because of a high gradient in the background, but by setting the maximum gradient to 5 (instead of the default value 3), all spots were accepted. One likely cause of the high background gradient for some of the spots was the very intense zero order reflection: the direct beam. Removing this reflection with a backstop was not possible without also removing all the reflections with a resolution lower than about  $5\text{\AA}$ . The reason for this is the low scattering angle and the position of the backstop in the CM-200 microscope. For protein crystallographic data collection the position of the backstop should have been much closer to the detector, but hardware restrictions prevented us from moving the backstop.

Scaling the data was problematic in view of the high partiality, the small wedge size the scarcity of data and the angular gaps between frames. For this purpose, we used the CCP4 program Scala (Evans, 2006). The statistics were extremely poor because we had to scale the partials using their calculated partiality and because of the large missing gaps of data between adjacent images. Figure 4.6 shows the merged and scaled 'h=0' zones of both data sets. Note that we processed all the data as P1 with anomalous signal. We did so in order to identify potential dynamic scattering, which results in a breakdown of point group symmetry in the diffraction patterns (Abrahams, 2010). The symmetries shown in figure 4.6 are therefore intrinsic to the diffraction experiment and not imposed by symmetry averaging of corresponding reflections.

---

## 4.4 Discussion

3D nano-crystals of lysozyme have successfully been grown, vitrified and transferred to an electron microscope at liquid nitrogen temperature. Rotation electron diffraction data was obtained at multiple places of a single sub-micrometer crystal to resolutions better than  $2\text{\AA}$ . For the first time, we could collect multiple frames of protein nano-crystals. This was made possible by the much higher sensitivity and signal-to-noise ratio of the Medipix-2 detector, combined with its radiation hardness at 200 keV.

By using a parallel electron beam, we could for the first time observe the curvature of the Ewald sphere in our electron diffraction patterns of protein crystals. At 200 keV, electrons have a wavelength of only about  $0.025\text{\AA}$ , causing the Ewald sphere to be extremely flat at low resolution. For instance, at  $5\text{\AA}$  resolution, the curvature of the Ewald sphere is less than  $0.15^\circ$ . The fact that in the diffraction images we can clearly see deviations in intensity between Friedel mates at  $5\text{\AA}$  resolution (figure 4.5), confirms the low mosaic spread suggested by the data processing. This implicates it should be possible to pinpoint areas of low mosaicity in the crystals. This may have an advantage if the crystalline order varies within a single nano-crystal, in which case its best parts can be selected.

These results are encouraging enough to consider fixing the technical problems that prevent optimal data collection. The lack of a scheduling interface between detector and the goniometer that controls the crystal orientation, combined with the relatively slow readout of the Medipix-2 quad (0.7s), is an evident cause of many of these problems. As the rotation of the sample could not be interrupted during the readout time, there were angular gaps between subsequent frames. We anticipate fixing this problem by employing a novel Medipix-2 detector on a RELAXd read-out board (Visser et al., 2011) with improved readout time (up to 200 frames per second) that can also read data while collecting the next frame. For small molecule crystals, it has been reported that collecting data in STEM-diffraction mode, further reduces beam damage (Kolb et al., 2011), so we plan incorporating the Medipix-2 detector in a microscope that has this option.

In view of the suboptimal data collection, we were happy being able to integrate the data with one of the standard programmes for processing X-ray data: MOSFLM. However, we cannot exclude the possibility that the results we report here do not reflect the true crystal structure. We noticed unexpected correlations between parameters. For instance, rotation of the crystal around  $\phi$  was accompanied by an apparent rotation (as suggested by post refinement) around  $\Psi_X$ , which is a rotation of the crystal around the beam. Perhaps this is caused by eucentric height variation of the crystal as it is being rotated. Since the electrons are focused spirally, rotation of the image or diffraction pattern with height is expected. These and other correlations, which were independent of the unit cell or crystal orientation parameters, may indicate that the indexing solution we present here is wrong. However, we think that at least in part, these correlations are caused by distortions specific for electron diffraction. For instance, the correlation of the beam shift with rotation of the crystal, which we observed (and corrected for), would not be encountered in X-ray diffraction.

Together with beam damage, the angular gaps between adjacent images may explain the observed breakdown in point symmetry observed in the 'h=0' Laue zones (figure 4.6). An alternative explanation may be found in the effects of dynamic scattering, which also causes such a breakdown in symmetry (Abrahams, 2010). With improved hardware we anticipate to be able to distinguish between these effects.

---

## 4.5 Conclusion

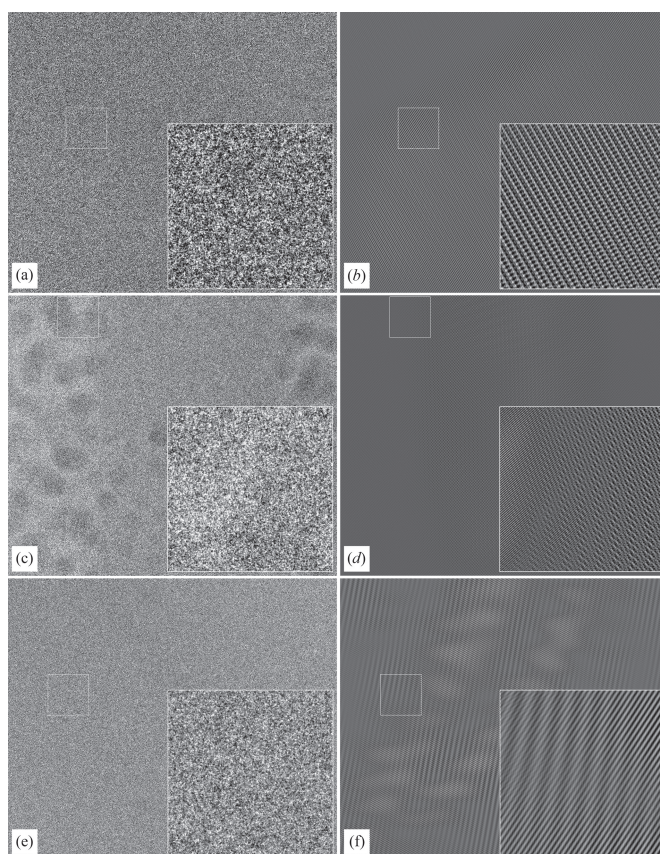
We have demonstrated that it is possible, at least in principle, to collect rotation electron diffraction data of protein nano-crystals. This prompts the question concerning the usefulness of such data sets for solving protein crystal structures. Problems with dynamic scattering have been anticipated, yet these did not prevent the structure solution to 3.3Å of the aquareovirus using single particle analysis (Zhang et al., 2010), which also has a size of approximately 100 nm, just like the crystals analyzed here. If at higher resolution, the effects of dynamical scattering cannot be ignored anymore, they can be modeled by the multi-slice approach (Jansen et al., 1998). This is common practice in electron crystallography of small molecules. Furthermore, there are theoretical considerations that imply that dynamical scattering may yield phase information (Evans, 2006), and a limited amount of such dynamical scatter may therefore even be beneficial.

Electron area detectors – like Medipix – that allow distinguishing the signal of high-energy electrons from that of other types of radiation are likely to have a major impact on structural biology. Our results imply that also protein crystallography stands to benefit from these technical advances.



## Chapter 5

### Lattice filter for processing image data of 3D protein nano-crystals



Eric van Genderen, Yao-Wang Li, Igor Nederlof and Jan Pieter Abrahams  
(Submitted, revision under review)

## Contents

<b>5.1 Abstract</b> .....	83
<b>5.2 Introduction</b> .....	83
5.2.1 2D versus 3D.....	84
5.2.2 Translational Symmetry.....	85
<b>5.3 Method</b> .....	86
5.3.1 Power Spectrum.....	87
5.3.2 Lattice filter .....	88
5.3.3 Detector artifacts .....	88
5.3.4 Data Acquisition.....	88
<b>5.4 Results</b> .....	90
5.4.1 Padding .....	90
5.4.2 Power Spectrum.....	90
5.4.3 Algorithm variables.....	92
<b>5.5 Discussion</b> .....	93
<b>5.6 Conclusion</b> .....	93

**Chapter image:** how beautiful are images taken with transmission electron microscopes: is it noise or signal? There can be structure in the noise! (Fig. 5.2)

---

---

## 5.1 Abstract

When 300 kV Cryo-EM images at Scherzer focus of  $\pm 100$  nm thick 3D protein nano-crystals are acquired by a Falcon 2 direct electron detector, Fourier transformation can reveal the crystalline lattice to surprisingly high resolutions, even though the images themselves seem devoid of any contrast. Here we report how this lattice information can be enhanced by means of a wave finder in combination with Wiener-type, maximum likelihood filtering. This procedure paves the way towards full 3D structure determination at high resolution of protein crystals.

## 5.2 Introduction

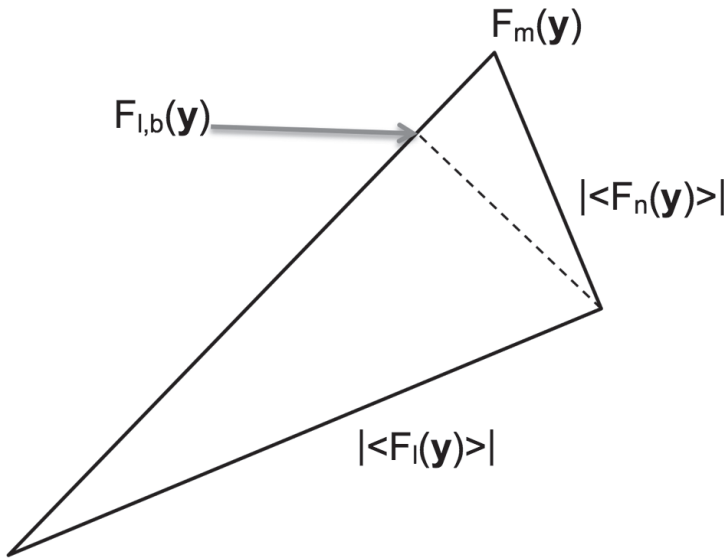
Previously, we demonstrated that electron diffraction of protein 3D nano-crystals could yield 2Å data (Chapter 4). Recently, we collected hundreds of high-resolution electron images at Scherzer focus of cryo-preserved, randomly orientated 3D nano-crystals of our test protein lysozyme (Nederlof et al., 2013a). Although appearing devoid of signal, Fourier transformation revealed crystalline order to a resolution of 4Å or better in about 50% of cases. Crystals with a thickness of about 100 nm (corresponding to 15 to 30 unit cells) yielded data with the best quality. The resolution of the Bragg spots in the Fourier transform of the electron micrograph is the lower threshold of the crystalline order. If the crystal occupies only part of the electron image, the remainder of the image contributes nothing but noise. If the crystal is cracked, twinned, warped, or contains mosaic blocks, the resolution of the Fourier transform is reduced because the unit cells do not align perfectly. In 2D crystallography the resolution is enhanced by computationally 'unbending' the crystal (Gil et al., 2006; Henderson et al., 1990; Kuhlbrandt and Wang, 1991). First, the 2D lattice repeat is identified and the frequencies not conforming to this repeat are filtered from the Fourier transform of the image by setting them to zero, thus enhancing the translationally repeating features of the image. This is equivalent to averaging the image with shifted versions of itself, whereby the magnitude and direction of the shifts are determined by the lattice parameters. This procedure will therefore average out noise that does not have translational symmetry, as it is not correlated to the signal. Hereafter the image of the crystal is subdivided into patches, which subsequently are aligned and averaged (Zeng et al., 2012; Stahlberg et al., 2001; Scherer et al., 2014).



### 5.2.1 2D versus 3D

One of the main differences between 2D crystals and 3D crystals, is that projection images of randomly oriented 3D crystals usually show moiré patterns, rather than regular 2D lattices. Hence it is not possible in the general case to extract a repeating unit as the third independent vector: defining the moiré usually is not a rational sum of the other two independent lattice vectors. The moiré pattern (a potentially non-repeating pattern that results from the superposition of multiple lattices) exists because of the three independent lattice parameters characterizing the 3D crystal. Due to Ewald sphere curvature, beam divergence and/or crystal mosaicity, repeats corresponding to each of these three translational symmetries can co-exist in the one and the same image. If one of the translation operators cannot be expressed as an integer sum of the other two, a moiré pattern results. While this pattern does not show directly the crystal lattice it does contain information of this a lattice.

Here we discuss a procedure for enhancing the moiré lattice information in the analysis of 3D nano-crystals, that does not require knowledge of the lattice parameters or orientation of the crystal; and even allows enhancing the lattices if multiple crystals are present in the image.



**Figure 5.1: Projection of  $|\langle F_l(\mathbf{y}) \rangle|$  onto  $F_m(\mathbf{y})$  to get the best estimate of the expected lattice structure factor  $F_{l,b}(\mathbf{y})$ .** The measured structure factor  $F_m(\mathbf{y})$  is the sum of the unknown structure factor corresponding to the lattice signal  $F_l(\mathbf{y})$  and the unknown structure factor corresponding to the noise  $F_n(\mathbf{y})$ . On average,  $F_l(\mathbf{y})$  and  $F_n(\mathbf{y})$  are uncorrelated, hence the expected absolute phase difference between them is  $\pi/2$ . We can only infer the absolute values of the expected structure factors. We cannot infer their phases, because there are two equally valid solutions mirrored over  $F_m(\mathbf{y})$ . The best estimate for  $F_l(\mathbf{y})$  is denoted by  $F_{l,b}(\mathbf{y})$ . It can be calculated by projecting  $|\langle F_l(\mathbf{y}) \rangle|$  onto  $F_m(\mathbf{y})$ .

### 5.2.2 Translational Symmetry

A high-resolution image of a (3D) crystal will have translational symmetry. Although this symmetry may not be obvious because of noise, the amplitudes of a Fourier transformation of the image will reveal the reciprocal lattice. One way of enhancing the translational symmetry in the image therefore is to identify the parameters that describe the reciprocal lattice, then to zero all reciprocal pixels that do not belong to this lattice and reverse the Fourier transform as in 2D crystallography. But this approach has a mathematical flaw, which intuitively can be understood as follows. Suppose a high background and a weak Bragg spot in the Fourier transform of the image. If all reciprocal pixels are set to zero, but the Bragg pixels are kept at their original value, then this will lead to a wrong estimation of the relative strength of this particular Bragg spot. This is equivalent to overweighting a weak Bragg spot relative to a stronger Bragg spot that is higher above the background. The Wiener filter addresses this problem in a maximum likelihood approach. We improve over existing methods by assuming that also the power spectrum of the noise and the lattice signal are uncorrelated. This is equivalent to establishing a Wiener filter that optimally enhances the lattice. Our approach does not need knowledge of the crystal lattice constants nor of crystal orientation or location to get a good result.

Assume a structure factor of the Fourier transform of the image:  $F_m(\mathbf{y})$ . It is the sum of the structure factor of the crystal lattice  $F_l(\mathbf{y})$  and the structure factor of noise  $F_n(\mathbf{y})$ :

$$F_m(\mathbf{y}) = F_l(\mathbf{y}) + F_n(\mathbf{y}) \quad (1)$$

Neither  $F_l(\mathbf{y})$  nor  $F_n(\mathbf{y})$  are known. We can only assume that they are uncorrelated. Their expected absolute phase difference will therefore be  $\pi/2$ , hence together with  $F_m(\mathbf{y})$  they define a right-angled triangle:

$$|F_m(\mathbf{y})|^2 = |\langle F_l(\mathbf{y}) \rangle|^2 + |\langle F_n(\mathbf{y}) \rangle|^2 \quad (2)$$

We can infer  $|\langle F_n(\mathbf{y}) \rangle|$  from the power spectrum of the image and use the result to calculate the expected amplitude of  $F_l(\mathbf{y})$ . However, we also require its phase and the only reasonable estimate is the phase of  $F_m(\mathbf{y})$ . So we need to project  $|\langle F_l(\mathbf{y}) \rangle|$  onto  $F_m(\mathbf{y})$  to get the best estimate of the expected lattice structure factor  $F_{l,b}(\mathbf{y})$  (Fig. 5.1).

Geometry implies the following equality, which is equivalent to an optimal (Wiener) filter (Press et al., 2007):

$$F_{l,b}(\mathbf{y}) = F_m(\mathbf{y}) \frac{|\langle F_l(\mathbf{y}) \rangle|^2}{|F_m(\mathbf{y})|^2} \quad (3)$$

Substituting with (2) gives:

$$F_{l,b}(\mathbf{y}) = F_m(\mathbf{y}) \left( 1 - \frac{|\langle F_n(\mathbf{y}) \rangle|^2}{|F_m(\mathbf{y})|^2} \right) \quad (4)$$

Thus scaling the structure factors of the original image by this likelihood, will recover the phases of the waves as well as their amplitudes. This theory leads to a more robust algorithm described below.

## 5.3 Method

In order to prevent wrap-around artifacts, we padded the images with pixel values which were set to the average value of the original image. The amount of padding can be defined by the user and corresponds to the expected size of the crystalline domains. The default value (used throughout the paper) is 1/16th of the image size (corresponding to 256 pixels for Falcon 2 images). The pixels of the Fourier transform  $F_m(\mathbf{y})$  of an electron image contain complex numbers (hence they carry phase information).

$$I(\mathbf{x}) = \mathfrak{F}^{-1}(F_m(\mathbf{y})) \quad (5)$$

### 5.3.1 Power Spectrum

First we calculate the radially averaged power spectrum  $|F_n(|\mathbf{y}|)|^2$  of the image in order to approximate the power spectrum of the noise:

$$|F_n(|\mathbf{y}|)|^2 = \int |F_m(|\mathbf{y}| \{ \cos \alpha \quad \sin \alpha \})|^2 d\alpha \quad (6)$$

The radial average (equation 5) is not completely smooth because of contributions of the spots at certain spacings. In order to correct for this, we assumed that the radial average of the 'noise' power spectrum is a decreasing function of  $r$ . So, if  $F_n(\mathbf{y})$  increases, this must be caused by the signal of Bragg spots. In that case, we keep  $F_n(\mathbf{y})$  constant until it drops below this value. An example of such a radial average is shown as a linear plot in figure 4, this is the result from the image in figure 2a. Then for each pixel, we calculate its significance  $s(\mathbf{y})$  as a normalized signal-to-noise ratio:

$$s(\mathbf{y}) = \frac{|F_m(\mathbf{y})|^2 - |F_n(|\mathbf{y}|)|^2}{|F_m(\mathbf{y})|^2} \quad (7)$$

In the absence of noise, ( $s(\mathbf{y}) = 1$ ) and when the norm of the signal is equal to the norm of the noise, ( $s(\mathbf{y}) = 0$ ). Note that ( $s(\mathbf{y}) = 1$ ) can be negative due to fluctuations in the noise level. (The real valued norm of a complex value is the square of its absolute value:  $|a + bi|^2 = a^2 + b^2$ ) In fact, in the absence of signal, fluctuations in the noise level will cause ( $s(\mathbf{y}) = 1$ ) to be negative for half of the reciprocal pixels! We consider a pixel to contribute significant information about the lattice, when  $s(\mathbf{y})$  is higher than a specified cutoff value. Pixels lower than the cutoff value are then set to zero. As a default, we used a cutoff value ( $c = 0.0$ ) for all the examples in this paper. So, for all examples given, all pixels of the lattice filter  $L(\mathbf{y})$  which had a norm below the radially averaged norm of  $F(\mathbf{y})$ , were set to zero. The value of the remaining pixels of the lattice filter  $L(\mathbf{y})$  were set to:

$$L(\mathbf{y}) = \frac{s(\mathbf{y}) - c}{1 - c} \quad (8)$$

### 5.3.2 Lattice filter

This lattice filter  $L(\mathbf{y})$  can be still be noisy, especially if the signal is low, so we included the option of only considering pixels likely to belong to a Bragg spot. The lattice parameters of the image are usually unknown at this stage of data analysis. So, for identifying potential Bragg spots, we used a method that does not require lattice parameters. First, we selected 3x3 clusters of pixels in which each of the pixels had a norm that is above a specified acceptance level  $a$ . We only allowed pixels to have non-zero values if they are less than a specified distance  $r$  away from any pixels within clusters which represents spots. This cut-off distance  $r$  is proportional to the reciprocal space equivalent of the expected size of the crystalline domains. All pixels of  $L(\mathbf{y})$  that are further away than  $r$  pixels from such a cluster of significant pixels (for which  $L(\mathbf{y}) > a$ ), were set to zero. As defaults we used an acceptance value ( $a = 0.4$ ) and a Bragg spot radius of ( $r = 4$ ) pixels for all the examples in this paper. This distance criterion can be suppressed by setting ( $a = c$ ).

### 5.3.3 Detector artifacts

Next to the lattice, this procedure also enhances other repetitive features of the image. Especially detector artifacts can be a major source of such spurious features. We found these artifacts to produce high resolution features. So we included an option to filter out such artificial signals by setting  $L(\mathbf{y})$  to zero for all ( $|\mathbf{y}| > n |\mathbf{y}|_{\max}$ ). As a default value we used a Nyquist cutoff of ( $n = 2/3$ ) for all examples in this paper.

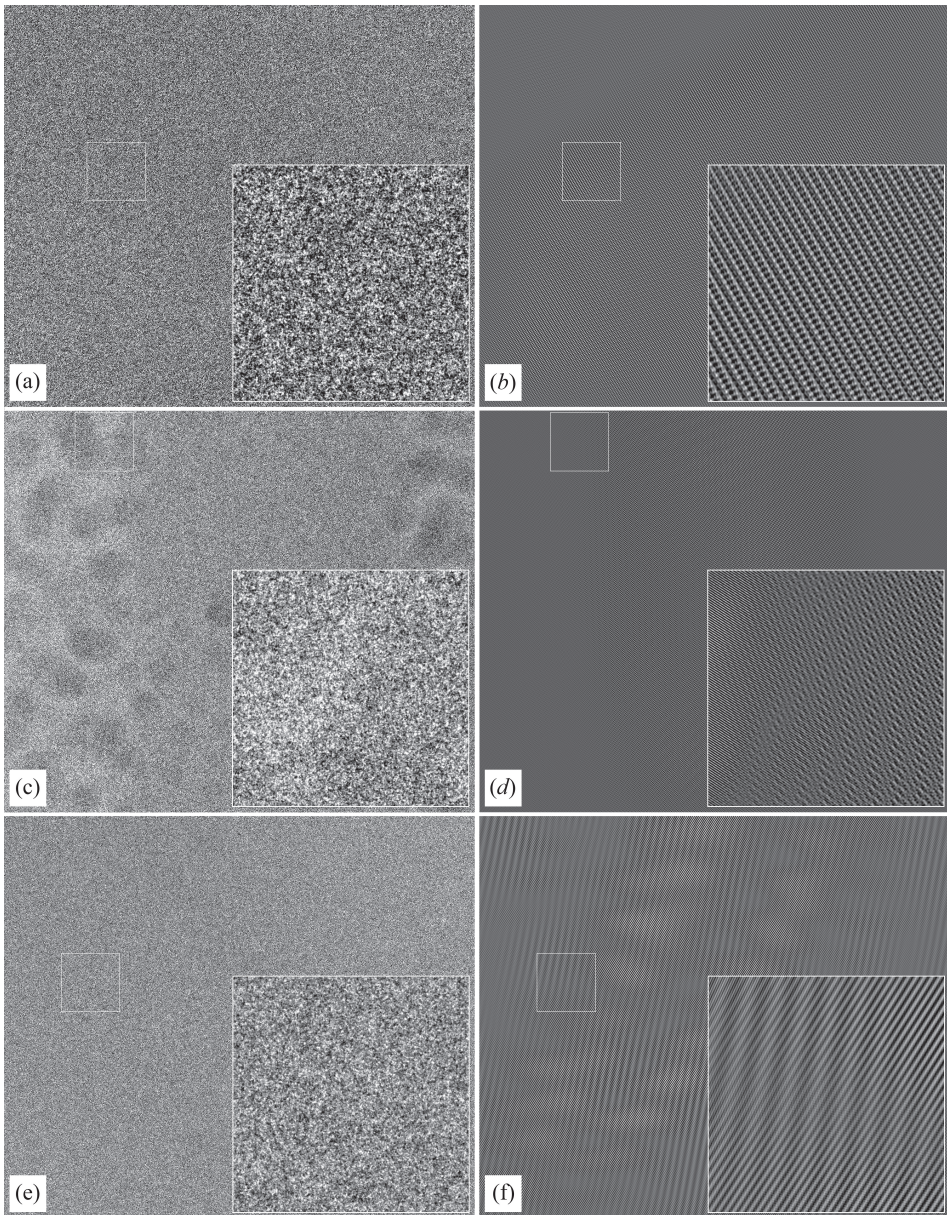
$$I_f(\mathbf{x}) = \mathfrak{F}^{-1}(L(\mathbf{y}) F(\mathbf{y})) \quad (9)$$

After having constructed the lattice filter  $L(\mathbf{y})$  according to the procedure outlined above, we calculated the filtered image  $I_f(\mathbf{x})$ :

### 5.3.4 Data Acquisition

The nano-protein crystal images used here as illustration of the method, were acquired on a FEI titan Krios (FEI, 2015) electron microscope at Scherzer focus of crystals with a thickness of approximately 100 nm. The data were collected with a Falcon2 FEI camera on 4048x4048 pixels at 0.5 s exposure time. The mean dose of the exposures was  $\sim 10 \text{ e}^-/\text{\AA}^2$  (for further details, see (Nederlof et al., 2013a)).





**Figure 5.2: Original images and filtered results.** 100 nm thick 3D crystals, produced by a Titan Krios 300 kV FEG transmission EM, captured with a Falcon 2 camera (4096x4096 pixels) using an 0.5 sec exposure time and an illumination of 3 to 10  $e^-/\text{\AA}^2$ . Left: original unprocessed images; right: processed images. (a, b) lysozyme crystal; (c, d) Lysozyme crystal with several ice crystals included (shown in detail); (e, f) crystal of a cross- $\beta$  peptide. Notice that in the cut-out of the results after filtering of middle image (d), we can distinguish an ice crystal (with a small unit cell and dominant spacings at  $\pm 3.8\text{\AA}$  at the left side of the cut-out) from the protein crystal (right side of the cut-out).

## 5.4 Results

Three example images (Fig. 5.2) from lysozyme nano-crystals show the merits of our new algorithm. The processed images show moiré patterns that are typical for non-oriented 3D crystals. Due to truncation errors, also some spurious repeating features will be visible in areas of the image where no crystal is present, but here the amplitudes are much lower than in the crystal. Thus the processed images will give clear indication where the crystal might be located. The result of the filter algorithm is shown in figure 5.2. This information can then be used for further analysis. (Nederlof et al., 2013b)

### 5.4.1 Padding

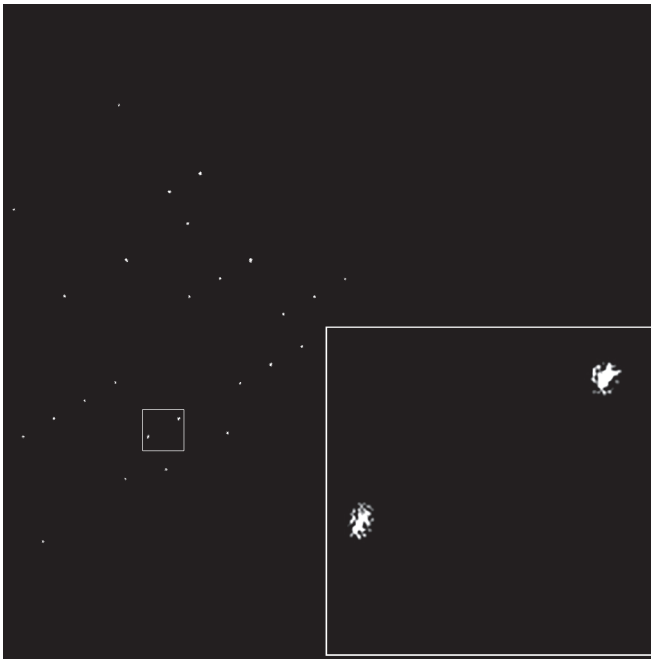
If the images are not padded as described in the Methods section, a wrap-around effect will occur and the information tends to bleed over the edge of the image into the opposite side of the image. This can be circumvented by padding the image, but when this padding is removed after lattice filtering, the resulting discontinuities at the image edges can lead to crosses centered on the Bragg spots in the reciprocal space, which could be undesirable for certain applications. Crosses can be prevented by choosing not to pad the images, or they can be suppressed by writing out filtered images without removing their padding (not shown).

### 5.4.2 Power Spectrum

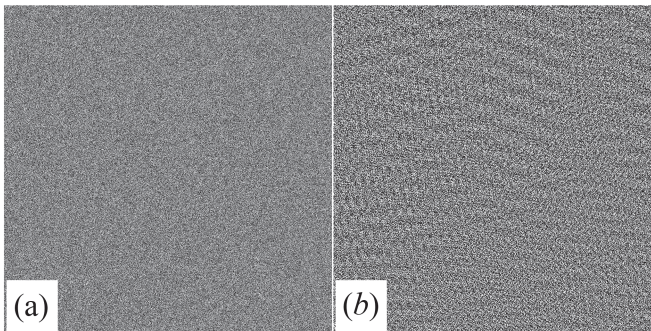
Close inspection of the power spectrum shows that the spots do not have a uniform shape (Fig. 5.3). The shapes of the spots differ when the projection of the crystal contains separate crystallites. The latter we call mosaicity and shows that within a crystal, domains of the crystal can exist and can each have a slightly different orientation. While this does not change the meta structure of the lattice, it will disrupt the moiré pattern and makes interpreting the structure more difficult. To put a positive twist on this: it also provides more orientations of the crystal.

How does our filter behave when applied to a image consisting of (generated noise) with the same median intensity and standard deviation as an image containing a crystal (values have been obtained from figure 2a)? The result of such a filtering is shown in figure 3. While some weak lattice structures can be seen, the amplitude of the wave structures is only slightly above the median pixel level. This shows that the filter is able to discriminate between an image consisting of random noise and a low S/N ratio image which contains a crystal lattice.

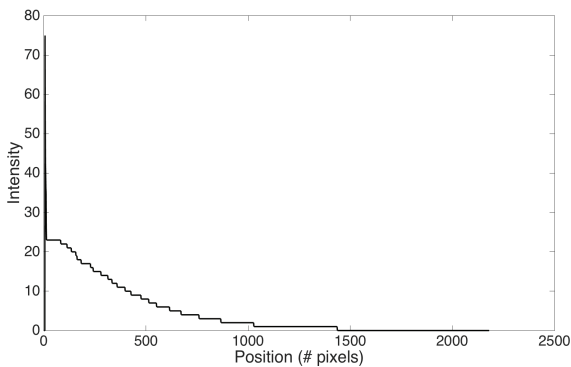




**Figure 5.3: Unique half of the centro-symmetric lattice filter of the lysozyme crystal image from figure 2.** Clearly visible are the spots that make up the lattice of the real image. Because of mosaicity and crystal shape, the spots are not uniform in shape.



**Figure 5.4: Filtered random noise.** (a) non-filtered random noise image with an intensity and standard deviation comparable to the image of a crystal in figure 2a (b) filtered random noise image

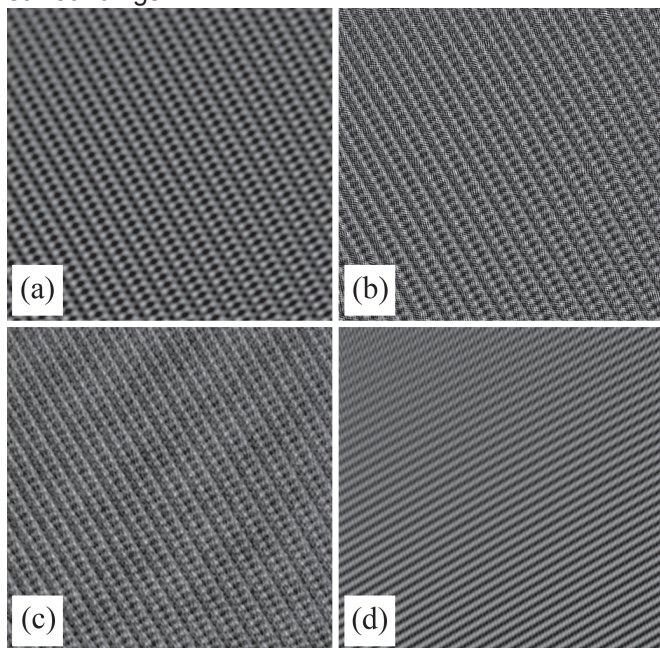


**Figure 5.5: Plot of the rotational average.** (figure 2c)



### 5.4.3 Algorithm variables

The algorithm includes the possibility to set a resolution cut-off corresponding to a factor of the Nyquist frequency. While this is a strong tool for removing certain detector artifacts or selecting a filter quality, misuse can introduce artifacts. Figure 6 demonstrates the effect of the lattice filter when using a sub-optimal Nyquist cutoff. (a) If it is too low, the high resolution spots will be obliterated and thereby the high resolution details of the crystal lattice will be lost. (b) When choosing the value to high this can result in a severe checkerboard striping and patterning due to detector artifacts. The artifacts introduced by a detector have structure-like features and will be of major influence in the final filtered image and should be taken into account. In the case of the Falcon 2 camera, used for the showcase sample images, this can include chip to chip fluctuations and read-out gain reference artifacts. The latter effects are most clearly visible at the Nyquist frequency in the power spectrum of the image. The Nyquist frequency artifact reveals itself as a bright line of pixels around the edge of the power spectrum, and a checkerboard or striped patterning over the whole filtered image. The other important filter parameter is the spot selection threshold. (c) If a value is chosen above the noise level, it will lead to selecting bogus Bragg spots. This will introduce spurious noise in the filtered image. (d) Another case is when criteria for selecting the spots too stringent and only a few of the brightest spots remain. In this case the final lattice image usually shows a 1D or 2D lattice pattern over the whole image, no longer discriminating between the crystal and its disordered surroundings.



**Figure 5.6: Filtering results from extreme variables.** Results from the same area as in figure 5.1. (a) Image with selecting a very small area of spots far from nyquist frequency. (b) Including the nyquist frequency. Visible are the artifacts from both including the camera as well as effects from wrapping effects in the reciprocal space. (c) Low spot threshold selection; by including spots that are not Bragg spots, more noise is present, but also more detail in the lattice is visible. (d) Very high spot selection threshold, thereby selecting only the strongest spots.

---

## 5.5 Discussion

In earlier work we reported interactive image processing to enhance the lattice. Though this produced good results, it was slow, tedious and required expert knowledge. Here we show the mathematical proof of our new approach, captured in an automatic algorithm that is very fast (half a second for a 4096x4096 image on a standard 2014 desktop computer). The lattice filter is a very powerful tool for selecting and analyzing extremely low contrast cryo images of 3D protein/peptide nano-crystals. It confirms that the 3D crystals are made up from multiple domains which are slightly differently oriented. Indeed, the algorithm can comfortably deal with multiple crystals with very different orientations, unit cells and/or space groups, as is witnessed in the middle panel of figure 2, which shows that the lattice of an ice crystal is enhanced just as well as the lattice of a protein crystal. Since more than two lattice parameters are required to describe the moiré lattice of a projected 3D crystal, using approaches from 2D crystallography cannot be applied straightforwardly or without considerable reprogramming. While one can argue that patches of the crystal with different orientation should not be back-transformed together, it is something that can be done directly after the filtering process (Nederlof et al., 2013a).

Our method does not correct for the CTF, but since it does not affect the phases of the projection image, a CTF correction can be done after lattice filtering. Although in principle a CTF correction could precede the lattice filter, it is advisable to first do the filtering, since this also takes care of the background removal. In the examples we give in our paper, we did not do any CTF corrections, as the data were collected at Scherzer focus, where the first sign reversal of the CTF occurs beyond the resolution limit of our images.

## 5.6 Conclusion

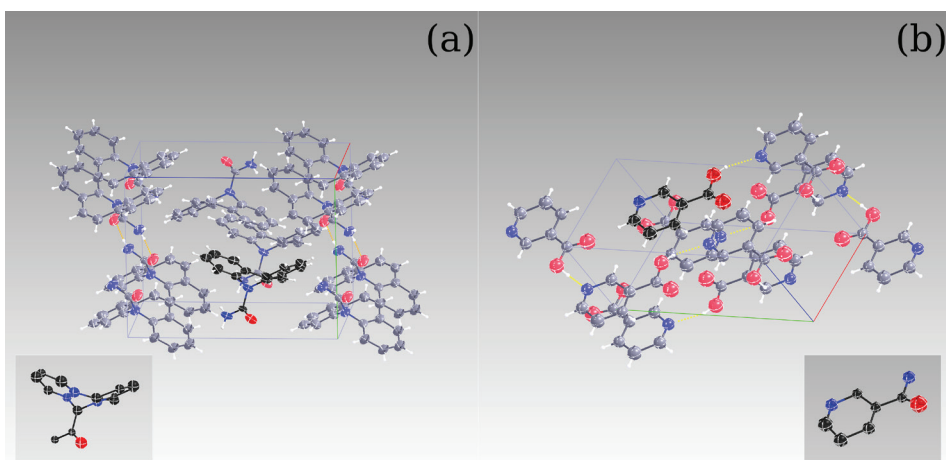
We propose the new lattice filter as a powerful tool for processing very noisy images with crystal structure factors (and thus with phase information) hidden within. The filter is able to discriminate between noise images and the very noisy images with very low contrast which contain crystal like structures. The lattice filter retains the shape of the spots in Fourier space and also retains any phase gradients within the Bragg spots (which determine the domain structure within the crystal). Thus it retains all the significant information of the Bragg spots. This will open the way to combine phases acquired by stationary, 2D images with intensities of rotation diffraction data taken from the same type of crystals. This way we expect to be able to phase the diffraction information of the protein and peptide crystals.

---



## Chapter 6

### ***Ab initio* structure determination of nanocrystals of organic pharmaceutical compounds by electron diffraction at room temperature using a Timepix quantum area direct electron detector**



Eric van Genderen, Max T.B. Clabbers, Partha Pratim Das, Tim Gruene, Andrew Stewart, Igor Nederlof, Kimberly C. Barentsen, Joaquim Portillo, Navraj S. Pannu, Stavros Nicolopoulos and Jan Pieter Abrahams  
***(Submitted, under revision)***

---

## Contents

<b>6.1 Abstract</b> .....	97
<b>6.2 Introduction</b> .....	97
6.2.1 Polymorphism.....	98
6.2.2 Detectors.....	99
6.2.3 Timepix detector .....	99
<b>6.3 Methods</b> .....	100
6.3.1 Camera.....	100
6.3.2 Microscope .....	102
6.3.3 Sample preparation .....	102
6.3.4 Data collection.....	103
6.3.5 Data pre-processing .....	103
6.3.6 Data processing .....	104
6.3.7 Structure solution & refinement.....	104
<b>6.4 Results</b> .....	106
6.4.1 Carbamazepine.....	107
6.4.2 Nicotinic acid .....	107
<b>6.5 Discussion</b> .....	108
<b>6.6 Conclusions</b> .....	109

**Chapter image:** after looking at the sparse data acquired, one can only wonder of the power of the methods used, to produce these results. (Fig. 6.4)

---

---

## 6.1 Abstract

Until recently, structure determination by transmission electron microscopy of beam sensitive 3D nano-crystals required electron diffraction tomography data collection at liquid nitrogen temperature, in order to reduce radiation damage. Here we show that the novel Timepix detector combines a high dynamic range with a very high signal to noise ratio and single electron sensitivity, allowing *ab initio* phasing of beam sensitive organic compounds, using low dose electron nano-diffraction ( $\approx 0.013\text{e}^-/\text{\AA}^2\text{s}$ ) data acquired at room temperature.

## 6.2 Introduction

Electron diffraction (ED) is a widely used tool for the determination of crystal structures of inorganic molecules and metals (Williams & Barry Carter, 2009). As electrons are less damaging than X-rays by several orders of magnitude relative to the amount of energy that is absorbed (Henderson, 1995), ED should be an attractive alternative to X-ray diffraction, when only nanometer sized crystals, short “nano-crystals”, of radiation sensitive molecules are available. However, difficulties with sample preparation and beam damage have remained major bottlenecks in transmission electron microscope (TEM) studies of pharmaceutical compounds and other beam sensitive materials. In X-ray crystallography, data are acquired by rotating the crystal during exposure about a constant axis normal to the direction of the beam. This method (known as the rotation method) has been the standard approach in protein X-ray crystallography for the last 40 years. It ensures good coverage of reciprocal space, it is experimentally less demanding than other approaches to data collection, it is optimal for radiation sensitive crystals, and it allows straightforward data processing (Arndt and Wonacott, 1977). The rotation method has more recently been adopted in the field of electron crystallography, where it has become known as “automated diffraction tomography” (ADT) (Kolb et al., 2008; Kolb et al., 2007; Mugnaioli et al., 2009). This and related methods like rotation electron diffraction (RED, (Wan et al., 2013; Yun et al., 2015; Zhang et al., 2010)) have been used to solve crystal structures of beam sensitive compounds *ab initio* from 3D ED data acquired at low electron dose from cryo-cooled samples.

## 6.2.1 Polymorphism

Pharmaceutical compounds can exist in different crystal packing forms called polymorphs. Different polymorphs can have different physical properties, like dissolution rate in water, stability, or bio-availability. Pharmaceutical compounds can often exist as multiple polymorphs and each one is separately patentable, so obtaining a polymorph portfolio of a specific pharmaceutical compound can also be of great economic value. To determine polymorphic crystal structures, both single crystal and powder X-ray diffraction can be used. For crystals of 0.5  $\mu\text{m}$  or larger, single crystal X-ray diffraction is adequate, but for smaller crystals beam damage becomes an issue. For such samples, X-ray powder diffraction is used. Since in X-ray powder diffraction the signal of a large number of crystals is averaged, the signal of individual nano-crystals will be lost in the signal of larger crystals. So this method may miss to detect one or more polymorphic modifications associated only with nano-crystals, especially if they are present in small amounts (e.g. less than 1-5 % wt). However, an ED study using a standard TEM may reveal the structure of any individual nano-crystal from 20 nm to several  $\mu\text{m}$  in size. So far, structure determination by ED of 3D nano-crystals of beam sensitive organic crystals has required data collection at liquid nitrogen temperature, in order to reduce radiation damage (Eddleston et al., 2010). There are multiple reasons to consider doing experiments at ambient temperature: (i) interference of ice crystals by ice formation during sample load; (ii) necessity of in-column cryo-plates to reduce contamination; (iii) possible change in crystal structure; (iv) reduced sample throughput, as loading/unloading samples can take up to 2 hours. There are examples of beam sensitive materials solved at ambient temperature by ED, but these structures were solved using the non-continuous ADT technique, e.g. Benzamides (Gorelik et al., 2012) and 9,9'-bianthracene-10-carbonitrile (CNBA) (Kolb et al., 2010).

---

## 6.2.2 Detectors

Hybrid pixel quantum area detectors (e.g. the Dectris single photon counting detectors (Brönnimann et al., 2002)) have revolutionized X-ray crystallography. The Medipix detectors that have been — and continue to be — developed by CERN— led consortia (Llopart et al., 2002; Llopart et al., 2007), have also been tested for electron detection (e.g. (McMullan et al., 2009; Georgieva et al., 2011)). They have much improved characteristics compared to CCD cameras often used to collect ED data. CCD cameras have several drawbacks for efficient ED data collection, namely: (i) blooming effects, where bright intense areas overshadow weaker ones located nearby; the need to use a beam stop to prevent damage to the camera; (iii) low signal to noise ratio (S/N); (iv) camera dark current; (v) increased background due to sensitivity to X-rays (abundantly produced in a TEM). Such drawbacks negatively influence the quality of ED data that can be collected from radiation sensitive samples, from organic pharmaceutical compounds to proteins.

## 6.2.3 Timepix detector

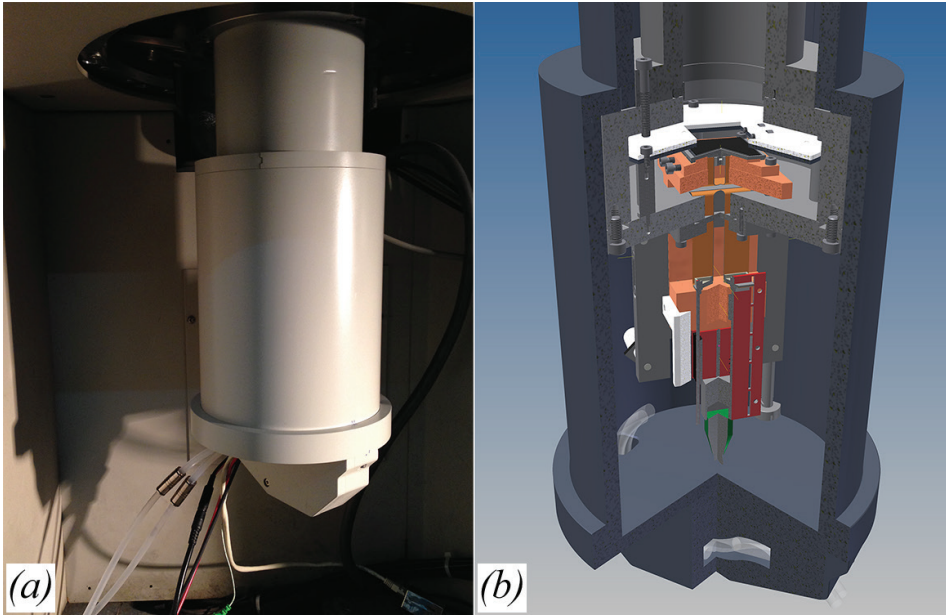
Here, we describe the benefits of the Timepix detector. The Timepix, which is a member of the Medipix family of detectors, is able to count only those particles that have an energy that is higher than a user-defined, programmable threshold. As the Timepix detector can discriminate between different types of quanta based on their energy, the noise signal from the abundantly present lower energy X-rays is filtered out. As a consequence, the detector has a very high S/N. Also dark current, which is present in CCD cameras due to thermal energy accumulation adding to static noise, is non-existent in the Timepix detector. At higher electron energies the corresponding scattering of the primary electron in the silicon detective layer can lead to charge sharing, resulting in counts in neighboring pixels other than the location of impact; this only leads to a slight spreading of the Bragg spots. Next to the very high S/N, the chip has a high linear range, a dynamic range of 11.8k electron counts with a linear range of 10k electron counts per acquired frame. Therefore very weak Bragg spots are visible next to very bright Bragg spots. These characteristics make the Timepix detector an ideal tool for quantitative diffraction. We investigate how Timepix would facilitate structure determination of radiation sensitive 3D nano-crystals without having to cryo-cool the sample.



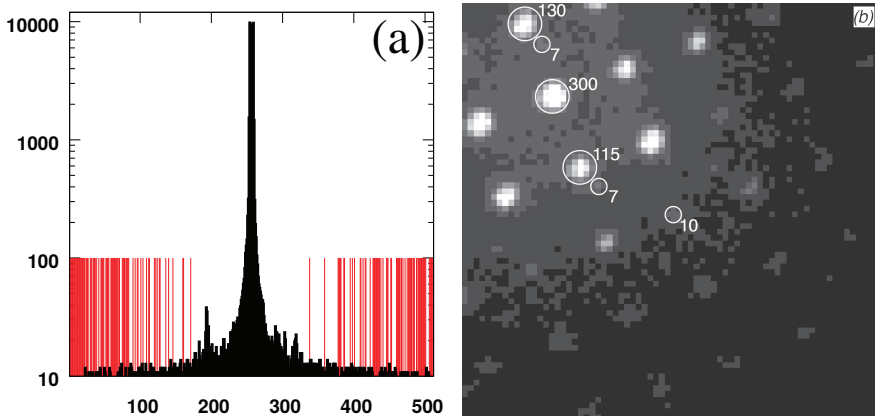
## 6.3 Methods

### 6.3.1 Camera

The camera we developed is modular and was fitted onto different TEMs, using different adapter flanges. (We speak of "camera" for the whole assembly of detector, read-out and pod parts. "Detector" is used for the Timepix chip including the sensitive layer). Up to now it has been tested on the following microscopes: Philips–FEI CM30 (LaB6), CM200 (FEG), CM300 (FEG), Zeiss Libra 120 (LaB6). The Timepix detector chip-assembly, its RelaxD readout electronics board and readout and control software were provided by Amsterdam Scientific Instruments (ASI, 2015). The Timepix chip was designed at CERN under the guidance of the Medipix collaboration (Llopart et al., 2007). The Timepix detector we used consists of 4 single Timepix chips which were bump-bonded to a Silicon sensor of 300  $\mu\text{m}$ . The thickness of 300  $\mu\text{m}$  is sufficient to capture all the energy of 200keV electrons while preventing damage to the Timepix ASIC (McMullan et al., 2009). A single Timepix chip has 256 by 256 pixels and a pixel size of 55  $\mu\text{m}$ , with larger pixels of 175  $\mu\text{m}$  on the edges. Therefore, a quad assembly has gaps between the individual 4 chips of twice 175  $\mu\text{m}$ , this leads to a higher count at pixels 256 and 257, which needs to be taken into account. Combining the 4 chips gives a total pixel count of 512 by 512 pixels. The detector was built into a vacuum pod and is actively cooled to  $20 \pm 0.1$  °C (Fig. 6.1). The RelaxD readout board, which can reach a frame rate of 120 fps (Visser et al., 2011), was outside the vacuum. The detector was calibrated using the SoPhy software from ASI, in order to ensure a uniform energy threshold for triggering each pixel. The cross section of an exemplary diffraction frame in figure 6.2a includes the direct beam position. It contains the dynamic range from zero counts to 10,020 on one frame. The edges contain continuous stretches of zero counts, i.e. the noise is due to scattering and the Timepix detector has no read out noise or dark current. The absence of "blooming" effects enables spot separation of weak peaks close to strong peaks, illustrated in figure 6.2b.



**Figure 6.1: “Laurel” Timepix detector.** (a) Detector mounted at the on-axis camera position on a CM30 (University of Barcelona, CciTUB services Spain). (b) 90° cut-off schematic of the camera as designed for the CM200 (Technische Universiteit Delft Netherlands).



**Figure 6.2: Timepix detector quality.** (a) Linear cross section through an image recorded with the Timepix detector ranges from 0 counts (marked red, set to 100 for visibility) to 10,020. Note the absence of a beam stop! (b) Spot separation of closely neighboring spots in the absence of “blooming” effects.

### 6.3.2 Microscope

A Philips CM30 LaB6 Microscope (University of Barcelona, CciTUB services) was operated in standard TEM mode at 200 kV. The very low dose rate was acquired by using the smallest available condenser aperture (20  $\mu\text{m}$ ), an extraction voltage step of 29 and spot size 9. This resulted in a dose of  $0.013\text{e}^-/\text{\AA}^2\text{s}$ . The beam was aligned to be as parallel as possible. This was achieved on the fluorescent screen at higher beam intensities, then slowly the spot size was lowered to 9 while keeping the beam shape and angle as constant as possible. (At the final steps the beam was acquired on the Timepix detector because the sensitivity of the fluorescent screen was no longer enough to visualize the electrons and thereby the image). The dose rate was measured by lowering the frame acquisition time to 0.1 ms. This way individual impacts were recorded; these were counted and the dose rate was interpolated. The single tilt rotation holder allowed rotating the stage from  $-60^\circ$  to  $60^\circ$ . The stage was tested and the tilt speed was linear at speeds below  $1^\circ/\text{s}$  for all angles. The semi-automatic alpha tilt stage required the operator to read the rotation angle directly from the stage providing a precision of  $0.3^\circ$ . The camera was operated with an “exposure time” of 0.1s, including dead time. This resulted in an acquisition rate of 9.1 frames per second (fps).

### 6.3.3 Sample preparation

The nano-crystalline samples were obtained by gently crushing dry crystals into a fine powder by rubbing them between two glass microscope slides. The powder was then carefully applied several times on a Holey Carbon 300 mesh microscope grid. The powder on different areas of the glass slides were applied on separated parts of the grid, in order to transfer a small amount of nano-crystals (from 50 to 500 nm in size) with random size distribution and orientation onto the electron microscope grid.

---

### 6.3.4 Data collection

Data were collected at 200 keV. Once a suitable crystal was found in TEM imaging mode at 5.3k magnification, it was centered by adjusting the Z height until the crystal remained at the same xy-position whilst tilting the holder. Data were collected by continuous rotation of the crystal (Arndt & Wonacott, 1977). The maximum tilt range was found manually by ensuring that the crystal remained in the beam over the entire observed tilt range. This range usually varied from 15° to 60°, depending on stage stability and the sample holder. For data collection, the microscope was switched to diffraction mode and the corresponding  $\alpha$  angle was recorded; when the maximum tilt angle was reached (where either the crystal was tilted out of the beam or the maximum tilt angle of the holder was reached) the corresponding final  $\alpha$  angle was recorded. ED frames were recorded during the stage rotation with the Timepix detector and data were saved as binary images to be further processed. Collection of a full rotation dataset required a few minutes (3–6 min) at most.

### 6.3.5 Data pre-processing

Datasets from rotation electron diffraction patterns were pre-processed for further analysis. Pre-processing (using in-house developed software) included: (i) correction for the spatial distortion and larger size of pixels at the horizontal and vertical central axes of the Timepix quad (ii) correct for any dead or bright pixels; (iii) centering of the images on the direct beam in case of high drift; (iv) summation of consecutive frames to ensure a suitable tilt step for data processing. After pre-processing, the images were transformed into PCK (Abrahams, 1993) for data processing using XDS.

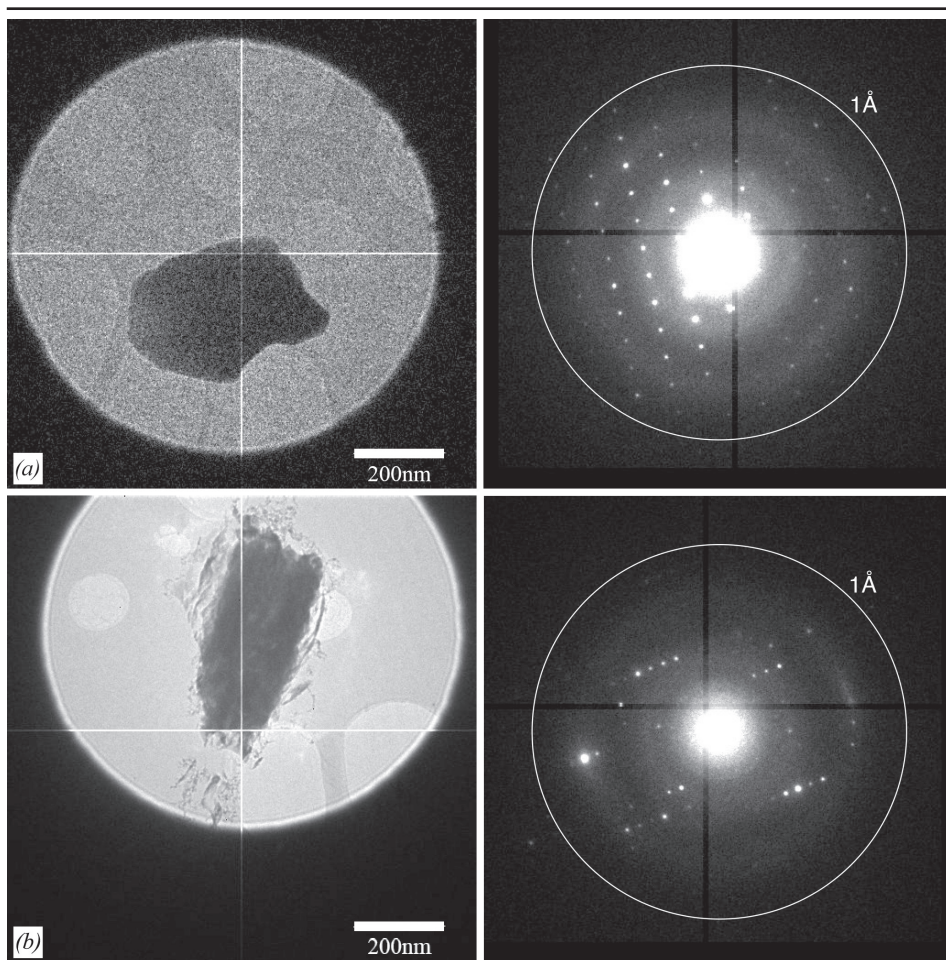
### 6.3.6 Data processing

Data were processed with XDS (Kabsch, 2010). For analysis with XDS, single frames were summed to create a data set with a rotational increment of approximately  $0.1^\circ$  per frame. The rotation axis was estimated from the diffraction image assuming that reflections with wide rocking curves are close to the rotation axis. The camera length was estimated from the powder diffraction of an evaporated aluminum standard grid with known d-spacings. The short wavelength of  $\lambda = 0.02508 \text{ \AA}$  leads to a very small  $2\theta_{\text{max}} \approx 1.76^\circ$  so that uncertainties in the camera length have a larger effect on the uncertainty of the unit cell parameters (see Fig. S1.1 Supplementary Information). In order to stabilize refinement of the experimental parameters, the Laue group was constrained to monoclinic, i.e.  $\alpha = \gamma = 90^\circ$ , and the camera length was not further refined. The high resolution cut-off was difficult to judge due to low completeness and low symmetry space groups. We cut the resolution very roughly at  $I/\sigma \approx 2$  and  $CC_{1/2} > 50\%$  (Karplus and Diederichs, 2012).

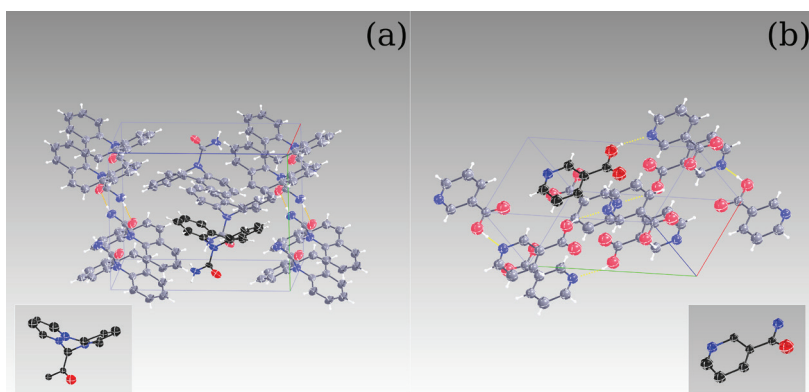
### 6.3.7 Structure solution & refinement

Data from XDS were prepared with XPREP (Bruker, 2004) for *ab initio* structure solution using the program SHELXT; subsequent refinement of the given structure was done with SHELXL and ShelXle (Sheldrick, 2008; Hübschle et al., 2011). The eight parameter fitting for electron scattering factors from (Peng, 1999) was used for SHELXL.

The carbamazepine structure was refined anisotropically unrestrained except for the RIGU restraint (Thorn et al., 2012). Nicotinic acid was refined isotropically without any restraints. Hydrogen atoms were added in riding positions. This results in 163 parameters for carbamazepine including the scaling factor between  $I_{\text{obs}}$  and  $I_{\text{calc}}$  with 144 restraints, and 39 parameters with no restraints for nicotinic acid.



**Figure 6.3: Rotation electron diffraction of pharmaceutical nanocrystals.** showing on the left panel an image of the crystal at 5.3k times magnification and on the right panel a diffraction pattern (a) a 200 nm thin carbamazepine crystal (b) a 200 nm thin nicotinic acid crystal



**Figure 6.4: Unit cell content for the final refined structures.** (a) Carbamazepine and (b) nicotinic acid for the single crystal data set, Table 1. Insets show the SHELXT solutions.



## 6.4 Results

Rotation electron diffraction data were collected at room temperature with a normal single tilt rotation holder of two different beam sensitive pharmaceutical compounds (carbamazepine and nicotinic acid), using a CM30 microscope. The CM30 had a modified flange which allowed fitting the housing of the Timepix detector in an on-axis position (Fig. 6.1). Stability of the goniometer and specimen holder limited the maximum range of continuous rotation at slightly more than 50° in some of the experiments. The Timepix detector was sufficiently sensitive to allow collecting high quality, low dose ED data of nano-crystals of these compounds for more than 5 minutes without any observable beam damage.

The data were of sufficiently high quality for both the structures of carbamazepine and nicotinic acid to be solved with direct methods. Given the uncertainty in the detector distance, the incomplete data and the temperature difference between the data presented here and literature values, our results as shown in Tables 1, (Supplementary Information S1.2–S1.6), including the remarkable agreement between our refined structures and published coordinates (El Hassan et al., 2013; Kutoglu and Scheringer, 1983). Figure 6.4 shows the unit cell content of the refined structures of carbamazepine (Fig. 6.4a) and nicotinic acid (Fig. 6.4b) against the single crystal data sets. The respective insets show the results from structure solution with SHELXT (see Supplemental Information for R1 values). Note that neither were additional atoms found, nor were any atoms missing after structure solution, and only a few atom types needed correction.

---

### 6.4.1 Carbamazepine

The carbamazepine nano-crystals were thinner than 200 nm during rotation and their electron diffraction patterns showed Bragg spots to 0.8Å resolution. In the low dose conditions that were used, the crystals diffracted at room temperature for more than 5 min to this resolution without visible deterioration. They could be kept within the beam for more than 50° of continuous rotation without needing intermediate realignment. Data were collected at a frame rate of 9.1 fps whilst maintaining a constant rotation speed. The total data acquisition time for any data set did not exceed 5 minutes (Table 1).

Data from a crystal of carbamazepine was collected with  $\Delta\phi = 0.018^\circ$  per frame and a total tilt of  $51^\circ$ . For data integration, six adjacent frames were summed, creating frames with an angular width of  $\Delta\phi_{\text{int}} = 0.108^\circ$  per frame. Indexing with XDS suggested a primitive monoclinic unit cell with axes  $a=7.53\text{\AA}$ ,  $b=11.14\text{\AA}$ ,  $c=14.06\text{\AA}$ , and  $\beta=92.80^\circ$  (Table 1). These parameters are consistent with carbamazepine polymorph III (El Hassan et al., 2013) measured by single crystal X-ray diffraction, which were:  $a=7.49\text{\AA}$ ,  $b=11.04\text{\AA}$ ,  $c=13.77\text{\AA}$ , and  $\beta=94.4^\circ$ , using a crystal with a volume that was 10 orders of magnitude larger than the crystal we used. With this single dataset we could already solve the structure with SHELXT (Sheldrick, 2015). Despite a light atom structure, SHELXT assigned all but three atom types correctly, did not miss any atoms and did not assign false peaks (see inset Fig. 6.4(a)). Subsequent structure refinement with SHELXL (Sheldrick, 2008) was stable even without any geometric restraints, as shown in Tables S1.3 and S1.4 (Supplementary Information).

### 6.4.2 Nicotinic acid

Continuous ED data were collected from a 200 nm thin nicotinic acid crystal. The whole tilt range spanned  $36^\circ$ , with the tilt range from  $-26^\circ$  to  $10^\circ$  with a frame width  $\Delta\phi = 0.048^\circ$ . For integration two frames were summed to  $\Delta\phi_{\text{int}} = 0.096^\circ$  per frame. Fig. 6.3b shows high resolution diffraction beyond  $1\text{\AA}$ . Indexing the diffraction patterns using XDS indicated a primitive monoclinic lattice and unit cell parameters consistent with those from X-ray data collected at 100K (Kutoglu & Scheringer, 1983) (see table 1). Tables S1.5 and S1.6 in the Supplemental Information summarize the good agreement between the structural model derived from these data and the published structure (Kutoglu & Scheringer, 1983).



## 6.5 Discussion

We collected and processed data from two small molecule compounds carbamazepine and nicotinic acid by electron diffraction from single nanometer-sized crystals at room temperature. We used a CMOS Timepix detector with a sensitivity about 10 times higher than conventional CCD cameras. This enabled data collection to high resolution of these radiation sensitive organic samples at very low dose without observable sample degradation ( $\approx 4.0\text{e}^-/\text{\AA}^2$  for the carbamazepine data set and  $\approx 1.1\text{e}^-/\text{\AA}^2$  for the nicotinic acid data set). Both data sets were of sufficient quality for structures solution by direct methods. To our knowledge this is the first time that crystal structures of organic nano-crystals could be solved with direct methods using ED rotation data of single nano-crystals collected at room temperature. The experimental limitation, due to the limited precision of the TEM specimen stage and the holder stability, resulted in very low data completeness, listed in table 1. Yet, in both cases unit cells parameters and the refined structures are consistent with literature values based on X-ray structure analysis. We show that our technique of ultra-fast 3D electron diffraction coupled with a sensitive Timepix detector allows fast and efficient 3D crystal structure analysis of organic pharmaceutical compounds at room temperature. The technique will allow higher throughput examination of nanometer size samples in a TEM at room temperature and can be complementary to the standard single crystal and powder X-ray diffraction. Future improvement of TEM goniometer stability for ED work and application of precession electron diffraction during data acquisition for precise atomic coordinates determination (Gemmi and Oleynikov, 2013) may extend the applications of electron crystallography to the wider TEM/X-ray scientific community. The techniques and methods shown here should pave the way to solve compounds with larger unit cell at non-cryo conditions and the possibility to solve the structures of even more beam sensitive protein crystals (Nederlof et al., 2013).

## 6.6 Conclusions

Electron diffraction data were recorded with the new Timepix CMOS detector in conventional rotation mode. The data quality was good enough to solve two organic crystal structures with direct methods. We describe the advantages of our approach, in particular the detector type that is well suited for the recording of electron diffraction data. Based on the experience from this project we are going to improve the instrumental design of both detector and diffraction instrument. We are also going to investigate how to produce better quality data. This will include processing programs dedicated to electron diffraction data and programs that take the effects of dynamic scattering into account.

**Table 6.1: Data statistics from XDS and model R-values after refinement with SHELXL.**

	Carbamazepine	Nicotinic acid
Molecular formula	$C_{15}H_{12}N_2O$	$C_6H_5NO_2$
Crystal size [ $\mu m^3$ ]	$1.2 \times 0.8 \times 0.2$	$0.7 \times 1.6 \times 0.2$
Tilt range [ $^\circ$ ]	51.00	36.00
$\Delta\phi$ [ $^\circ$ /frame]	0.018	0.048
$\Delta\phi_{Int}^\dagger$ [ $^\circ$ /frame]	0.108	0.096
Dose [ $e^-/\text{\AA}^2$ ]	4.0	1.1
Space group	P2 <sub>1</sub> /n	P2 <sub>1</sub> /c
Cell <sup>‡</sup>		
$a, b, c$ [ $\text{\AA}$ ]	7.487(1), 11.041(2), 13.775(3)	7.186(2), 11.688(3), 7.231(2)
$\beta$ [ $^\circ$ ]	92.94(4)	113.55(6)
Cell dimensions		
$a, b, c$ [ $\text{\AA}$ ]	7.53(1), 11.139(6), 14.06(2)	7.30(1), 11.693(2), 7.33(3)
$\beta$ [ $^\circ$ ]	92.80(8)	113.7(1)
Resolution [ $\text{\AA}$ ]	8.73-0.8 (0.85-0.80)	5.82-0.75 (0.86-0.75)
$R_{merge}^\S$ [%]	8.4 (35.8)	7.1 (34.9)
$I/\sigma$	5.64 (1.80)	4.96 (1.75)
CC <sub>1/2</sub> <sup>¶</sup>	99.0 (53.0)	99.4 (83.1)
Completeness (%)	45.0 (46.2)	35.6 (36.0)
Reflections	2202 (371)	953 (152)
Unique reflections	1071 (181)	503 (82)
$R_{complete}^\parallel$	32.2	37.9
R1 (%)	28.0	35.6
wR2 (%)	55.6	63.9

<sup>†</sup> frames were summed before integration

<sup>‡</sup> (El Hassan *et al.*, 2013) for Carbamazepine and (Kutoglu & Scheringer, 1983) for for Nicotinic Acid

<sup>§</sup>  $R_{merge} \equiv R_{sym} \equiv R_{int} = \sum_h \sum_j \frac{|I_{hj} - \langle I_h \rangle|}{\langle I_h \rangle}$

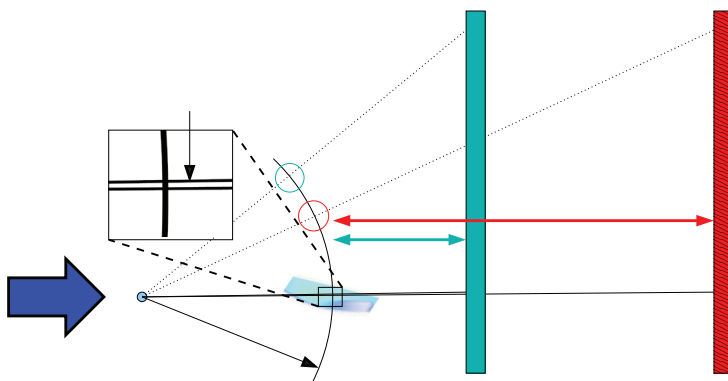
<sup>¶</sup> (Karplus & Diederichs, 2012)

<sup>||</sup> (Luebben & Gruene, 2015)



## Supplementary Information

***Ab initio* structure determination of nanocrystals of organic pharmaceutical compounds by electron diffraction at room temperature using a Timepix quantum area direct electron detector**



**Chapter image:** It requires a special kind of thinking to explain a difficult concept with a few simple lines. (Fig. 6.S1)

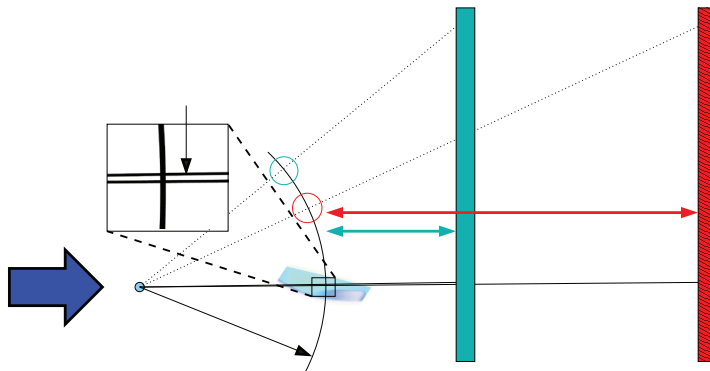


Figure S1: A short wavelength leads to greater errors in the unit cell parameters. At short wavelengths typical tiucro electron diffraction,  $\lambda = 0.0251\text{\AA}$ , the maximum diffraction angle  $2\theta$  is very small. See text for details.

## S1 Camera Length and Cell Parameters

The wavelength used for electron diffraction is typically about  $\lambda \approx 0.025\text{\AA}$ , whereas in X-ray diffraction, it is about  $\lambda \approx 0.5\text{--}1.5\text{\AA}$ . The short wavelength leads to a small maximum diffraction angle,  $2\theta_{\max} \approx 1\text{--}2^\circ$ , as opposed to  $2\theta_{\max} \approx 40\text{--}170^\circ$  in X-ray diffraction. In order to calculate the reciprocal lattice, the recorded diffractions are projected onto the Ewald sphere. At short wavelength, even a threefold greater camera lengths leads to only small variations of the lattice. Therefore, uncertainties in the detector distance lead to uncertainties in the unit cell parameters.

## S2 Structure Solution

Data processed with XDS were solved with SHELXT. As we could not use the default options, we list the options for each data set here together with a summary of the results:

### Carbamazepine, Single Crystal (Table 1)

```
$ shelxt cbz_a005 -a -z1.5 -m1000 -k10 -q0 -s"P2(1)_n"
  R1  Rweak Alpha  Space group  Flack_x  Formula
  0.352 0.296 0.281  P2(1)/n      C14 N3 O
```

### Carbamazepine, Merged Data from five crystals (Table S1)

```
$ shelxt cbz_mrg -a -z1.5 -m1000 -k10 -q0 -s"P2(1)_n"
```

---

R1	Rweak	Alpha	Space group	Flack_x	Formula
0.452	0.248	0.422	P2(1)/n		C11 N6 O2

**Nicotinic Acid (Table 1)**

```
$ shelxt na_b004 -a -m1000
```

R1	Rweak	Alpha	Space group	Flack_x	Formula
0.378	0.078	0.150	P2(1)/c		C6 N2 O
0.382	0.047	0.096	Pc	-4.21	C7 N O4
0.346	0.057	0.134	P2(1)	-2.29	C11 N O4

**Nicotinic Acid, Merged data from two crystals (Table S2)**

```
$ shelxt na_rg -a -m1000
```

R1	Rweak	Alpha	Space group	Flack_x	Formula
0.460	0.390	0.184	P2(1)/c		C5 N3 O
0.410	0.167	0.068	P2(1)	no Fp	C10 N2 O7
0.441	0.198	0.220	Pc	no Fp	C9 N4 O3

**S3 Merging Data**

The main article shows data statistics for single data sets from carbamazepine and from nicotinic acid. We collected additional data sets from five different crystals for carbamazepine and two different crystals for nicotinic acid in order to see how merging affects the data. The data are shown in Table S1 and Table S2 respectively. The merging statistics in both cases are worse than for the single data sets, and structure solution with *SHELXT* required more options (see Sec. S2). It is not clear whether this is due to the effect of dynamic scattering. The amide group of carbamazepine can be oriented in two ways. During structure refinement against a data set of a single crystal we could observe a difference in *R1* between both possibilities, so that one crystal appears to crystallise with a preferred orientation. Several crystals may be a mixture of either orientation which could be a reason for the poor statistics. However, this is only a hypothesis. As this question is not the central question of this work, we did not investigate the effect at this stage.

Data were merged using *POINTLESS/AIMLESS* [1] as merging with *XSCALE* produced worse results. We are currently investigating the reasons. Note that *AIMLESS* has more strict rejection criteria than *XSCALE*, hence the sum of reflections in Table S4 from each crystal is greater than the total number of reflections for the merged data set. *POINTLESS* calculates a weighted averaged cell from the unit cell parameters.

Table S1: Data processing statistics of XDS from five different carbamazepine crystals and  $R$ -values from refinement with SHELXL. Res: resolution range, high resolution shell in brackets where applicable. Compl: % Completeness. Refl: Total number of reflections.

	Crystal 1	Crystal 2	Crystal 3	Crystal 4	Crystal 5	5x merged
$\phi_{\text{total}} [^\circ]$	50	38.8	37.1	40	41	
$\Delta\phi$	0.039	0.016	0.012	0.022	0.031	
$\Delta\phi_{\text{Int}}^1$	0.117	0.098	0.098	0.088	0.093	
Cell						
$a$	7.58	7.58	7.58	7.55	7.59	7.58
$b$	11.17	11.20	11.19	11.17	11.21	11.20
$c$	13.99	13.93	13.93	13.96	14.03	13.93
$\beta$	93.1	92.7	93.0	92.7	93.4	92.7
Res. [Å]	8.73.0–0.81 (0.87–0.81)	11.20–0.81 (0.86–0.81)	8.72–0.91 (0.96–0.91)	7.54–0.83 (0.88–0.83)	11.21–1.16 (1.23–1.16)	11.19 – 0.81 (0.87–0.81)
$R_{\text{merge}} [\%]$	6.7 (37.2)	8.3 (36.4)	8.1 (38.5)	7.0 (22.1)	21.9 (42.0)	16.2 (50.8)
$I/\sigma$	6.18 (1.74)	7.05 (1.79)	5.91 (1.80)	7.96 (1.84)	3.09(1.99)	7.0 (1.8)
$CC_{1/2}$	99.4 (62.2)	99.5 (76.5)	99.3 (69.3)	99.4 (82.5)	95.7 (63.1)	99.0 (84.3)
Compl	45.1 (48.6)	26.2 (27.0)	28.9 (26.9)	17.9(3.7)	27.4 (29.4)	77.5 (62.3)
Refl.	1950	1685	1364	1402	1617	5415 (456)
Unique	1066	607	668	654	615	1861 (237)
$R_{\text{complete}}$						45.6
$R1 [\%]$						43.2
$wR2 [\%]$						71.1

Table S2: Data processing statistics of XDS from two different nicotinic acid crystals and  $R$ -values from refinement with SHELXL. Data for “Crystal 1” identical with Nicotinic Acid entry in Table 1. n/a: not applicable; n/c: not computed

	Crystal 1	Crystal 2	2x merged
Tilt range [°]	36.0	36.2	n/a
$\Delta\phi$ [°/frame]	0.048	0.021	n/a
$\Delta\phi_{\text{Int}}^2$ [°/frame]	0.096	0.107	n/a
Cell			
$a$	7.30(1)	7.42(1)	7.406
$b$	11.693(2)	11.690(7)	11.692
$c$	7.33(3)	7.509(7)	7.377
$\beta$	113.7(1)	115.21(5)	114.45
Resolution [Å]	5.82–0.75 (0.86–0.75)	11.70–1.16 (1.23–1.16)	11.69–0.75 (0.84–0.75)
$R_{\text{merge}}$ (%)	7.1 (34.9)	11.2 (34.8)	21.0 (32.9)
$I/\sigma$	4.96 (1.75)	3.90 (1.85)	4.7 (1.9)
$CC_{1/2}$	99.4 (83.1)	98.9 (92.7)	97.6 (85.9)
Completeness (%)	35.6 (36.0)	34.8 (29.1)	38.6 (37.9)
Reflections	953 (152)	304 (31)	1266 (274)
Unique	503 (82)	139 (16)	542 (147)
$R_{\text{complete}}$	37.9	n/c	44.2
$R1$ [%]	35.6	n/c	40.2
$wR2$ [%]	63.9	n/c	69.7



## S4 Model Quality

In order to assess the data quality beyond the statistical descriptors in Tables 1 and 2, a comparison of the bond lengths and angles for Carbamazepine with the values published by [2] is shown in Tables S3 and S4. All structures were refined without restraints. Standard uncertainties as provided by SHELXL.

Table S3: Bond lengths for carbamazepine in comparison with the high resolution structure by [2]. *single*: single crystal data set. Table 1; *5x merged*: merged data from five crystals, Table S1. Lengths listed in [Å].

	<i>El Hassan</i>	<i>single</i>	<i>5x merged</i>
N1-C15	1.3815(5)	1.42(3)	1.37(2)
N2-C15	1.3566(4)	1.37(3)	1.39(2)
C15-O1	1.2384(4)	1.19(3)	1.24(2)
N1-C1	1.4302(4)	1.40(4)	1.45(2)
N1-C14	1.4316(4)	1.41(3)	1.43(2)
C9-C14	1.4041(4)	1.41(2)	1.42(3)
C1-C6	1.4054(4)	1.49(3)	1.38(3)
C8-C9	1.4624(5)	1.45(4)	1.45(2)
C6-C7	1.4619(5)	1.55(3)	1.45(3)
C7-C8	1.3520(6)	1.24(4)	1.34(3)
C5-C6	1.4080(4)	1.37(4)	1.35(3)
C4-C5	1.3883(5)	1.38(4)	1.43(3)
C3-C4	1.3966(6)	1.33(3)	1.36(3)
C2-C3	1.3930(6)	1.46(4)	1.38(3)
C1-C2	1.3959(5)	1.36(2)	1.44(3)
C9-C10	1.4063(5)	1.46(3)	1.40(3)
C10-C11	1.3882(6)	1.36(5)	1.35(3)
C11-C12	1.3959(6)	1.33(3)	1.41(3)
C12-C13	1.3936(6)	1.42(3)	1.38(3)
C13-C14	1.3988(5)	1.49(4)	1.43(2)

Table S4: Selected bond angles for carbamazepine in comparison with the high resolution structure by [2]. Angles in  $^{\circ}$ . Column headers as in Table S3.

	<i>El Hassan</i>	<i>single</i>	<i>5x merged</i>
C1-N1-C14	117.13(3)	119(1)	116(1)
O1-C15-N2	122.26(3)	122(1)	120(1)
C7-C8-C9	126.25(3)	128(2)	129(2)
C6-C7-C8	127.75(3)	126(2)	126(2)
C8-C9-C14	123.04(3)	124(2)	124(2)
C1-C6-C7	123.44(3)	119(2)	125(2)

Table S5: Bond lengths of nicotinic acid in comparison with the high resolution structure by [3]. *single*: single crystal data set. Table 1; *2x merged*: merged data from two crystals, Table S2. Lengths listed in  $\text{\AA}$ .

	<i>Kutoglu</i>	<i>single</i>	<i>2x merged</i>
N-C	1.348(4)	1.36(2)	1.41(3)
C1-C2	1.397(4)	1.41(1)	1.46(3)
C2-C3	1.406(4)	1.44(4)	1.50(5)
C3-C4	1.383(5)	1.42(3)	1.33(3)
C4-C5	1.393(5)	1.41(3)	1.36(4)
C5-N	1.342(4)	1.28(4)	1.26(5)
C2-C6	1.490(4)	1.46(2)	1.53(3)
C6-O1	1.308(4)	1.34(3)	1.32(5)
C6-O2	1.211(4)	1.16(2)	1.14(3)

Table S6: Selected bond angles for nicotinic acid in comparison with the high resolution structure by [3]. Angles in  $^{\circ}$ . Column headers as in Table S5.

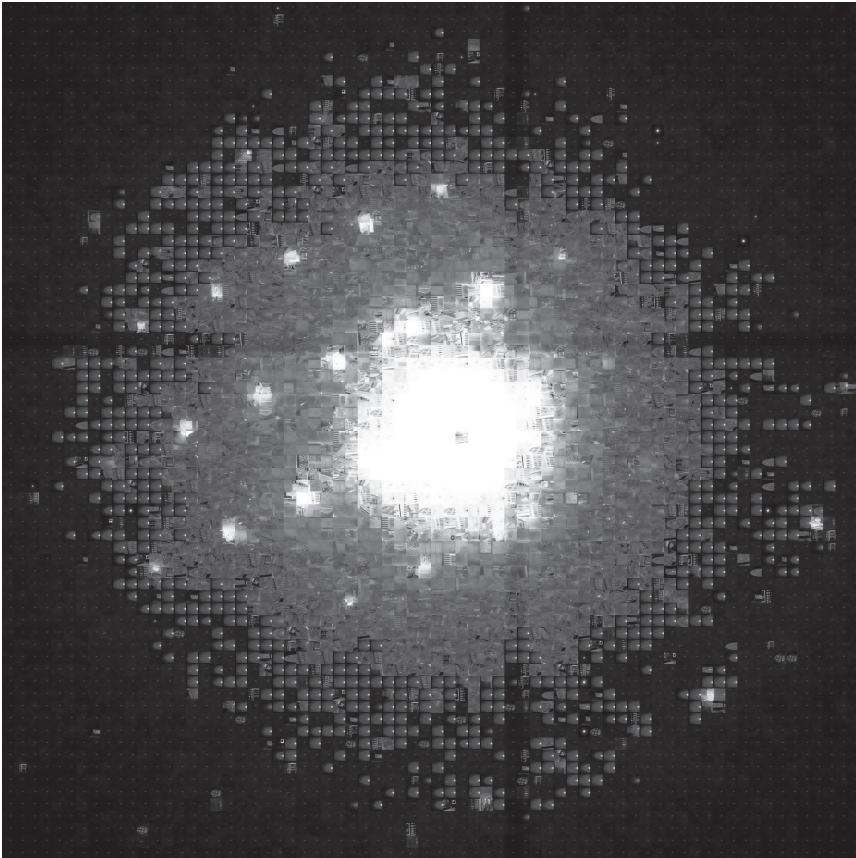
iucr

	<i>Kutoglu</i>	<i>single</i>	<i>2x merged</i>
C1-N1-C14	117.13(3)	119(1)	116(1)
O1-C15-N2	122.26(3)	122(1)	120(1)
C7-C8-C9	126.25(3)	128(2)	129(2)
C6-C7-C8	127.75(3)	126(2)	126(2)
C8-C9-C14	123.04(3)	124(2)	124(2)
C1-C6-C7	123.44(3)	119(2)	125(2)



## Chapter 7

### General discussion and Future prospects



## Contents

<b>7.1 introduction .....</b>	<b>121</b>
<b>7.2 Crystallisation of proteins .....</b>	<b>121</b>
7.2.1 Crystal growth .....	121
7.2.2 Practice .....	122
<b>7.3 Detector and camera development .....</b>	<b>122</b>
7.3.1 CMOS based monolithic electron detectors .....	122
7.3.2 Medipix & Timepix hybrid pixel detectors .....	123
<b>7.4 Protein (lysozyme) crystallography.....</b>	<b>123</b>
<b>7.5 Phasing of 3D nano-crystals by lattice filtering .....</b>	<b>124</b>
<b>7.6 <i>Ab initio</i> structure determination of organics .....</b>	<b>125</b>
<b>7.7 Concluding remark.....</b>	<b>125</b>
<b>7.8 Future prospects .....</b>	<b>126</b>
7.8.1 TEM.....	126
7.8.2 Phasing .....	126
7.8.3 Solution .....	127

**Chapter image:** artist impression; to look into the future is learning from the past. Image build from two diffraction datasets and all photos taken during the my PhD.

---

---

## 7.1 Introduction

In the past decade, advances in structure determination with electron microscopy of organic, beam sensitive, materials have been significant. The newly developed techniques, triggered by new microscope systems (like the FEI Titan Krios systems (FEI, 2015)) and new cameras, made it possible to acquire 3D structural information from these samples to a resolution which was impossible to achieve before. Knowledge is required to improve structure solution and every aspect of the process involved, from treatment of radiation sensitive materials, sample preparation, TEM imaging and diffraction systems all the way to how data must be interpreted. In this thesis I explained multiple new techniques and methods developed by us, using both new microscopes as well as a new type of detector: Timepix. I describe how these tools can help to overcome (what were) the most important problems and bottlenecks in detection of very low dose ED.

## 7.2 Crystallization of proteins

For the development of new methods and techniques a good sample is essential. Without a good, known sample, it is impossible to quantify and understand the quality of any new method. Every research field has their own standard for method development, for example: zebrafish in biology, *E.coli* in microbiology and mice in medical research; in (X-ray, Electron) crystallography lysozyme for protein crystallography and carbamazepine for small organics are used. Lysozyme has a well known characterized crystal structure with over 100 polymorphs reported and the crystallization properties for different sizes and shapes of crystals are well known. While this was the protein of choice in our studies, other proteins will require different crystallization conditions and handling

### 7.2.1 Crystal growth

Crystals go through three phases: nucleation, growth and cessation of growth. For crystallization the protein needs to be pure and conformationally uniform. The trick is to reach supersaturation and to find the best crystallization strategy to get the protein through the different phases; to create the meta-stable thermodynamical properties which are required. First the type of protein which is crystallized is to be considered and initial screenings can be based upon this. Crystallization techniques which currently exist are: vapour diffusion, batch crystallization, gel crystallization and dialysis. The work of crystallization screening is very laborious; therefore multiple automations are used for mass screening of conditions. This includes robots which can do multiple gradient essays and automated imaging systems.

## 7.2.2 Practice

When good crystals are found it is good practice to test if they contain proteins. Multiple methods exist and among them are: *in situ* diffraction analysis, using UV to detect the fluorescence signal of tryptophan and second harmonic microscopy (SONICC). In this thesis, I describe how the methods from chapter 3 are used to produce nano-crystals of for example lysozyme, of sufficient thickness to take ED (chapter 4). New techniques are, and will be, developed to make the production of crystals from specific types of proteins more convenient.

## 7.3 Detector and camera development

Up until ten years ago the most optimal detectors for imaging and diffraction experiments in TEM were based on the same principle. Especially for radiation hard materials, the amount of radiation, and therefore the data acquisition quality of the detector was not an issue. Film or image plates were used if high resolution and dynamic range was needed but acquisition speed was not important. CCDs were used if speed was required. However, in the case of CCD cameras, dark current as well as the noise from X-rays (which also introduce signal that cannot be separated from the signal of the primary electrons,) made them less suitable for tomography experiments of radiation sensitive materials like organic compounds and zeolites.

### 7.3.1 CMOS based monolithic electron detectors

Recently, multiple cameras have been developed to deal with this problem in the case of imaging. Their main characteristics are: a high number of very small pixels and a method for very precise electron position detection. Examples are the FEI Falcon 2 (FEI, 2015) and Gatan K2 (Gatan, 2015) cameras. Both are CMOS based monolithic electron detectors. By making their detective layer underneath the integrating circuitry as thin as possible, they ensure to have a very precise electron detection at high electron energies above 200 KeV. However this means that at high dose the electronics directly above the sensitive layer will slowly get damaged. Also the dynamic range is very limited. The very aim of these detectors is to maximize DQE and minimize the MTF by back thinning the sensitive layer; however, this makes them highly radiation sensitive. A diffraction with intense spots will inevitably damage the electronics of these type of detectors. Because of these, for diffraction experiments, major drawbacks we used a detector which combines radiation hardness with a very good S/N ratio.

---

### 7.3.2 Medipix & Timepix hybrid pixel detectors

The Medipix family of hybrid pixel detectors supports a thick sensitive Si layer which captures the full energy of the incident electron. The total energy deposited in each pixel is then compared to a fixed, but programmable, energy threshold. If this energy is above the threshold, in CM mode it is then counted as a hit. The Timepix detector has the ability to also work in a ToT mode to measure the total energy deposited in a pixel as a function of time pulses. If this can be combined with the moment of arrival (ToA) of the energy packets, one could have a better understanding of the point of impact and thereby improving the detector characteristics (DQE and MTF). However this would require the newer Timepix-3 to fully exploit this mode of data acquisition.

The camera designs supporting these hybrid pixel detectors have been improved significantly. In chapter 2 multiple iterations of the camera are shown together with a description of the improvements on: cooling, shielding and support. A proper cooling is important for detector stability. We decided to use a Peltier element with a temperature feedback-loop from a position near the detector, because it was easy to implement and use. We placed the read-out electronics outside the vacuum but inside the lead mantle to have easy access. The connection between the detector and read-out consisted of a flex cable through a custom vacuum feed-through. With the new designs, data of high quality could be measured from crystals substantially smaller than the wavelength of visible light: nano-crystals.

## 7.4 Protein (lysozyme) crystallography

In chapter 4 I describe how data can be taken of 3D Lysozyme protein crystals at multiple angles; the data shown in this thesis was acquired on a Philips/FEI CM200 TEM with a special cryo-stage (TU Delft, The Netherlands). As a detector we used a Medipix 2.1 Quad detector. This detector was able to take frames at a base speed of ~1 fps. We were able to collect semi-continuous rotation datasets. This proved that if better detectors and hardware would be available, it would be possible to take full continuous rotation datasets of protein nano-crystals (as described in Chapter 6).



## **7.5 Phasing of 3D nano-crystals by lattice filtering**

It might be possible to get good phase information from imaging 3D nano-crystals (Nederlof et al., 2013). With this information it should be possible to phase diffraction data acquired with TEM. To prepare the data for further analysis it was not possible to use standard computational filters available for 2D crystal analysis; this because multiple shifted layers of proteins in the crystal which are superimposed upon each other will create moiré patterns.

In Chapter 5 I describe how such a filter can be enhanced and automated using a very fast algorithm. It contains improvements for the step-by-step method described in the original paper (Nederlof et al., 2013) by introducing certain thresholds and solutions for artefacts (artefacts originating from image transformation as well as the camera). This lattice filter is very robust and is able to enhance lattices of protein crystals, but just as well of ice crystals captured together with the protein crystal.

---

## 7.6 *Ab initio* structure determination of organics

I describe in chapter 6 how the continuous rotation method as suggested in chapter 4 combined with a better and much faster Timepix detector, can be used to collect very good data. We were able to collect, rotation sets of sufficient quality to solve *ab initio* the structure of sub-micrometer crystals of small organic pharmaceutical compounds by direct methods at ambient temperatures ( $\sim 30^{\circ}\text{C}$ ). The data had a resolution up to  $0.8\text{ \AA}$  for the full rotating of up to  $51^{\circ}$  for a total time of 3-5 min. This resulted in a total dose of on the crystal of  $\sim 1.3\text{ e}^{-}/\text{\AA}^2$ .

To validate the data XDS was used. We found the space group of carbamazepine to be  $P2_1/n$  with the unit cell parameters  $a=7.53\text{ \AA}$ ,  $b=11.14\text{ \AA}$ ,  $c=14.06\text{ \AA}$ , and  $\beta=92.8^{\circ}$ . We found the space group of Nicotinic Acid to be  $P2_1/c$  with the unit cell parameters  $a=7.30\text{ \AA}$ ,  $b=11.69\text{ \AA}$ ,  $c=7.33\text{ \AA}$ , and  $\beta=113.7^{\circ}$ . These values are close to the unit cells found by X-ray crystallography. With better equipment it should be possible to improve the unit cell to match the one from literature since these datasets were still very incomplete. The results prove that it is possible to investigate very small crystals and solve their structure *ab initio*. This should pave the way to: (i) solve compounds with larger unit cell at ambient temperature conditions; (ii) solve structures from even smaller crystals (iii) to use this improved method for protein crystals under cryo conditions.

## 7.7 Concluding remark

The development of diffraction techniques and methods is very promising. With diffraction it is possible to get more high resolution information at more angles from a certain area of interest than when imaging techniques are used. This is due to the shot noise (electrons which did not interact with the sample). These electrons are distributed over the whole image ( $\sim 90\%$  of the total electrons). This is different from diffraction where these electrons are contained in the (low resolution) central beam of the diffraction pattern. This really becomes clear when comparing the data from chapter 4 with the paper of Nederlof et al (2013) where the crystals were imaged at multiple angles; it is possible to take 400 diffraction patterns with the Timepix camera of different rotation angles, where in imaging only up to two images could be acquired.

## 7.8 Future prospects

The results presented and their conclusions in this thesis point out multiple aspects of electron nano-crystallography that could be improved. As shown the camera is no longer the main bottleneck (Chapter 6). is to consider the drawbacks of conventional TEMs: (i) the stage is usually unstable over long ranges and if rotation data of more than  $50^\circ$  is to be acquired, this should be improved; (ii) also a TEM is a compromise between imaging and diffraction. While not described in this thesis, we already obtained data of higher resolution using a FEI Titan Krios than with other (older) TEM, however due a static lens problem there is a linear elliptical distortion in the magnification. This can be corrected in imaging mode, however it will stay in diffraction mode. Also the Cs corrector in a TEM required for high resolution imaging is difficult to use and not necessary for diffraction.

### 7.8.1 TEM

Another thing that can be improved is sample preparation. A better method to distribute the protein nano-crystals on the grid. Some of the problems are: (i) big crystals that coat a large area with a thick ice layer, hiding smaller crystals but also increasing dynamical scattering; (ii) too many crystals on top of each other, also resulting in thick ice layers but also in interfering diffraction patterns. (iii) water depleted protein crystals, where the water layer was completely removed. A possible solution would be spraying the crystals on known positions and freeze them instantaneously while they are still encapsulated by a small layer of water.

### 7.8.2 Phasing

The phasing problem diffraction faces could be solved in multiple ways: (i) by imaging the 3D crystals of the proteins and using the phases measured to phase the information of diffraction taken at the same angles. (It would be beneficial if the diffraction pattern was taken at the same position and rotation angle as the image.); (ii) use the dynamic scattering in the sample and use the differences in intensity created by the dynamical scattering between the Friedel mates as a measure (Abrahams, 2010). (iii) another, completely novel way would be to shape the beam into a known form with a known phase shift, and then use this information to phase the diffraction pattern.

---

### 7.8.3 Solution

All problems and solutions described above could be tackled with the design of a dedicated electron diffraction device. This would be a more radical solution than taking an existing TEM and modifying it.. However designing a new type of device would have its merits: (i) when designed with a horizontal electron beam, the whole device can be as large as a standard X-ray beam line; (ii) a dichromatic beam could be implemented. Using electrons at two different energies / wavelength and after the beam has been transmitted through the sample it will be split into an imaging and diffracting beam as a solution for the phase problem. Designing such a device would remove major bottlenecks in the acquisition of good diffraction data and will change 3D nano-crystallography of very beam sensitive materials. If we even look beyond diffraction of crystalline samples, such a device might even make it possible to do diffraction of non crystalline samples.



## Summery

This thesis is focused on different steps or improvement on the data collection and structure solution of beam sensitive organic 3D crystals, like pharmaceuticals and proteins. We show, that with the use of a novel detector like the Timepix and its predecessor the Medipix-2 big improvement can be made to data collection of these crystals. Also we describe improvements on some existing algorithms to recover the phases of diffraction patters.

**Chapter 1** describes some common concepts in diffraction, crystallography and data collection of beam sensitive crystals. This includes transmission electron microscopes (TEMs), electron scattering, imaging, diffraction and detection methods. It also contains a short introduction to different technologies available in 2D and 3D crystallography, including X-ray, transmission electron microscopy and X-ray free electron lasers (FEL). Also is discussed how these technologies and their methods support each other as dependent methods within their common niches. Each part in this chapter explains its merits and limitations of the different methods, explaining why data collection required novel approaches. The chapter is concluded with a short thesis outline.

In **chapter 2** a brief overview is given about the different detector (Medipix) variants available. This is complemented with the characteristics of the Timepix detector, which was the latest iteration of the Medipix family we used. I explain the concept of data quantum efficiency (DQE) and modulation transfer function (MTF) and explain their shortcomings when applied as a quality descriptor for hybrid pixel detectors, like the Timepix, when used for diffraction experiments. The second part of the chapter describes the three cameras which were designed for these experiments. Briefly I discuss the requirements for each camera and how we included them in the camera design (e.g. cooling, vacuum and radiation shielding.)

In **chapter 3** we cover some of the main aspects of protein nucleation and crystallization. Different diagnostic tools, crystallization techniques and strategies are explained. New tendencies in the field such as combining heterogeneous nucleants and high throughput methods are also presented.

In **chapter 4** we report on data collection on a cryo-TEM combined with a Medipix-2 detector of lysozyme 3D nano-crystals. Protein crystallography is a major justification for large-scale X-ray facilities like synchrotrons and free electron lasers. Protein crystals smaller than 0.5  $\mu\text{m}$  are too small for X-ray crystallography and 30% of the proteins, that crystallize do not produce crystals bigger than that; however, to study these crystals we can exploit the characteristics of electrons. Since electrons are a couple of magnitudes less damaging than X-rays per diffracted quantum we show in this chapter that good data can be obtained using the Medipix-2 detector before the crystal is destroyed. For the first time it was possible (in principle) to collect rotation sets of a couple of degrees, with, because of the excellent properties of the Medipix-2 detector, resolutions up to 3Å.

In diffraction all phase information is lost. If the resolution of a diffraction pattern is above 1.1Å the structure can not be solved ab initio by direct methods. One of the possibilities to add phase information to a dataset is to use images of the same type of crystal (or even better the same crystal at the same location). In an earlier paper we concluded that phase information could be recovered (Nederlof et al., 2013a), however the algorithm used to enhance the lattice was very slow and not well explained. In **chapter 5** we show improvements on the algorithm and explain in detail its merits for enhancing lattice information of 3D crystals. Because of the shot noise the images to be enhanced were very noisy; also the images were taken at Scherzer focus and therefore they seemed to be devoid of any signal. The improvements to the lattice filter include better spot selection, where shape of the spots is retained in Fourier space and a way to retain any phase gradients within the Bragg spots. Thus, it retains all the significant information of the Bragg spots. This will open the way to combine phases acquired by stationary, 2D images with intensities of rotation diffraction data taken from the same type of crystals.

---

To test the Timepix detector for diffraction data collection in very low electron flux/dose conditions we pushed the boundaries of continuous rotation electron nano-crystallography. Using the pharmaceutical compounds carbamazopine and nicotinic acid we took continuous rotation diffraction datasets at ambient temperature conditions. In **chapter 6** we show that with a Timepix detector it was possible to gain information of up to 0.8Å resolution. With sufficient degrees of rotation it was possible to solve the structure of these small compounds ab initio by direct methods without any prior information. We also explain why the camera is able to detect information at these super low flux conditions. e.g. The very high S/N ratio combined with its high dynamic range and radiation hardness gave excellent data.

Finally we conclude in this chapter that with the experience from this project we are going to improve the instrumental design of both detector and diffraction instrument. We are also going to investigate how to produce better quality data. This will include processing programs dedicated to electron diffraction data and programs that take the effects of dynamic scattering into account. Furthermore we will investigate the use of this same method (at cryo-conditions) for much bigger organic compounds of up to 1 kDa and for proteins.

Finally in **chapter 7** we explain how each of the steps taken, as described in this thesis, improved our understanding of electron nano-crystallography. Especially, we emphasize the Timepix detector from the Medipix family-type of detectors. The use of such a dedicated camera is a very valuable tool in very low dose experiments. We conclude that without such a detector it would be almost impossible to collect such quality data. Also we draw general conclusions about gaining phase information for these structures, as when data is acquired lower  $\sim 1.0$  Å it is not possible to solve the structure without.

In the future prospects we present some solutions for the problems we faced. We suggest how improvements on the TEM itself and the quality of the sample used, as these are now the main bottlenecks of the method, will be the next steps to take. However, we also suggest to not neglect the search for even better cameras with better characteristics.





---

## Samenvatting

Dit proefschrift beschrijft verscheidene stappen en verbeteringen op het gebied van data collectie en 3D structuur oplossing van straling gevoelige organische 3D kristallen, waaronder farmaceutische moleculen en eiwitten. We laten aan de hand van de Timepix detector en zijn voorganger de Medipix-2 detector zien dat grote verbeteringen met betrekking tot data collectie mogelijk zijn. Ook beschrijven wij verschillende verbeteringen van enkele bestaande algoritmen om de fase te herleiden van de fase-loze diffractie patronen.

**Hoofdstuk 1** behandelt enkele gebruikelijke concepten op het gebied van diffractie, kristallografie en data collectie van straling gevoelige kristallen. Dit hoofdstuk bevat onder andere basis informatie over transmissie elektronen microscopen, verstrooiing van elektronen, het opnemen van data in zowel beeld- als diffractiemodus als wel verschillende opname technieken. Het bevat ook een korte introductie tot de verschillende technologieën die momenteel beschikbaar zijn op het gebied van 2D en 3D kristallografie, waaronder X-ray, transmissie elektronen microscopie en X-ray vrije elektronen lasers. Ook beschrijf ik hoe deze technologieën en hun methoden elkaar kunnen ondersteunen op de punten waar de de ander tekort schiet. In elk gedeelte van dit hoofdstuk beschouwen we de sterke en zwakke punten van de verschillende methoden en verklaren we waarom nieuwe benaderingen nodig zijn op het gebied van data collectie. Het hoofdstuk wordt vervolgens afgesloten met een kort overzicht van de verschillende onderdelen die het proefschrift bevat.

In **hoofdstuk 2** geef ik een kort overzicht van de verschillende detector varianten die beschikbaar zijn van het Medipix type; aangevuld met de verschillende karakteristieken van de Timepix detector. De Timepix detector van de Medipix-familie is het laatste type dat we gebruikt hebben. Ik beschrijf de concepten achter data quantum efficiëntie (DQE), de modulatie overdracht functie (MTF) en geef commentaar over hun tekortkomingen om de kwaliteit van hybride pixel detectoren te beschrijven wanneer zij worden toegepast worden bij diffractie experimenten. Het tweede gedeelte van dit hoofdstuk gaat in op het inbouwen van deze detectoren in behuizingen voor TEM. Met name laat ik zien hoe wij de verschillende vooraf gestelde eisen voor het design hebben toegepast (bijvoorbeeld op het gebied van vacuum, koeling en stralingsbescherming).

In **hoofdstuk 3** beschouwen we een aantal van de belangrijkste aspecten op het gebied van nucleatie en kristallisatie van eiwitten. Verscheidene diagnostische hulpmiddelen, kristallisatie technieken en strategieën worden uitgelegd. Nieuwe ontwikkelingen op dit gebied worden ook besproken, zoals bijvoorbeeld het combineren van heterogene nucleanten en high throughput methoden.

In **hoofdstuk 4** rapporteren wij over de data collectie van lysozym 3D nano-kristallen op een cryogene-TEM met behulp van de Medipix-2 detector. Een groot deel van het bestaansrecht van X-ray faciliteiten zoals synchrotrons en vrije elektronen lasers komt voort uit de wens om de 3D structuur van eiwitten te kunnen bepalen. Kristallen kleiner dan 0.5 micrometer zijn te klein voor X-ray studies en 30% van de eiwitten die kristalliseren, produceren geen kristallen groter dan dat; echter, voor dit type kristallen is het mogelijk om ze te bestuderen door gebruik te maken van de verschillende karakteristieken van elektronen. Omdat elektronen veel minder schade aanrichten dan Röntgenstralen per quantum dat gediffracteed wordt, laten we zien dat kwalitatief goede diffractie data kan worden verkregen door gebruik te maken van de Medipix-2 detector, voordat het kristal teveel schade oploopt. Hier laten we zien dat met de excellente eigenschappen van de Medipix-2 het mogelijk was om rotatie series op te nemen van een aantal graden met resoluties tot aan 3Å

Alle fase informatie gaat verloren bij een diffractie van een kristal. Als de resolutie niet hoger is dan 1.1Å is het niet mogelijk de structuur *ab initio* op te lossen met behulp van directe methoden. Één van de mogelijkheden om de fase weer toe te voegen aan de informatie verkregen uit diffractie is om beelden te gebruiken die zijn opgenomen in beeldmodus van de TEM van eenzelfde type kristal; met voorkeur voor hetzelfde kristal en hetzelfde gebied op dit kristal. Al eerder hebben wij beschreven dat fase informatie uit de beelden van 3D kristallen kan worden gedestilleerd (Nederlof et al., 2013a). echter dit algoritme was hier niet goed beschreven en erg langzaam om uit te voeren. In **hoofdstuk 5** laten we verbeteringen zien met betrekking tot dit algoritme en gaan we in detail in op de toegevoegde waarde voor het beter zichtbaar maken van het kristalrooster van 3D kristallen. Door de ruis die veroorzaakt wordt door “shot noise” en omdat de beelden worden opgenomen bij Sherzer focus lijken ze totaal geen informatie te bevatten. De verbeteringen voor het filter bevatten onder andere: (i) een nieuwe manier om Bragg spots te herkennen; (ii) een manier om de fase gradiënten binnen de Bragg spot te behouden. Door deze twee vernieuwingen te combineren, kan de random noise worden genegeerd, maar alle significante informatie in de Bragg spots behouden blijven. Deze methode maakt het mogelijk om de fases uit stationaire 2D beelden te combineren met de intensiteiten van rotatie series in diffractiemodus van hetzelfde type kristallen.

Om de Timepix detector te testen in super lage elektronen flux hebben we de huidige regels en grenzen omtrent wat mogelijk was met lage elektronen flux experimenten moeten verleggen. We maakten gebruik van organische moleculen, welke in de farmaceutische industrie worden toegepast (carbamazopine en nicotine), om in omgevingstemperatuur condities, continue rotatie series op te nemen in diffractiemodus. In **hoofdstuk 6** laten we zien dat met behulp van de Timepix detector het mogelijk was om een data resolutie te behalen van  $0.8\text{\AA}$ . Door voldoende rotatie graden te maken was het voor deze kleine moleculen mogelijk om de structuur direct *ab initio* op te lossen zonder voorkennis van de structurele informatie. Ook maken we een korte beschrijving waarom deze detector het mogelijk maakt om in diffractiemodus bij deze lage flux toch kwalitatief goede data te verzamelen. o.a. De hoge signaal tot ruis verhouding samen met het hoge dynamische bereik en stralingshardheid van deze detector maakte excellente data mogelijk. Ter afsluiting maken we in dit hoofdstuk de belofte dat we, met behulp van de in dit project opgedane kennis, het instrumentarium van zowel de detector als het diffractie instrument verder te verbeteren en te ontwikkelen. We gaan ook onderzoeken hoe nog kwalitatief betere data kan worden opgenomen. Dit moet worden behaald, onder andere, door het aanpassen en ontwikkelen van programma's speciaal voor elektronen diffractie data en programma's die het dynamisch verstrooien van elektronen meeneemt in hun evaluatie, maar bijvoorbeeld, ook deze verstrooiing gebruikt voor verbetering van de structurele informatie. Verder gaan we kijken of deze methode ook toepasbaar is voor veel grotere organische moleculen ( $\sim 1\text{ kDa}$ ) en eiwitten.

Tot slot bespreken we in **hoofdstuk 7** hoe met elke stap die genomen is onze kennis van elektronen nano-kristallografie heeft vergroot. De nadruk ligt vooral op hoe de camera en met name de Timepix heeft bijgedragen aan het verkrijgen van goede data van kristallen met onder een zeer lage flux van elektronen. Wij stellen in dit hoofdstuk dat zonder een detector met de eigenschappen van de Timepix, het bijna onmogelijk is om opnamen te maken van deze uitmuntende kwaliteit. Ook benadrukken wij het belang van een goede methode om de fase van de structuren te kunnen herleiden.

We presenteren aan het einde van het hoofdstuk ook nog ideeën en oplossingen voor de problemen geschetst hierboven, om in de toekomst nader te bestuderen en te onderzoeken. We geven aan dat, hoewel het onderzoek naar nieuwe camera's en detectoren door moet gaan, dit niet langer de beperkende factor is. Het onderzoek zal zich in de nabije toekomst vooral moeten richten op de verbetering en ontwikkeling van nieuwe diffractie instrumentaria en nieuwe methoden om het monster te prepareren.

## References

- [Abrahams, 1993] Abrahams, J. P. (1993). Compression of x-ray images. *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.*, 28:3–4.
- [Abrahams, 2010] Abrahams, J. P. (2010). The strong phase object approximation may allow extending crystallographic phases of dynamical electron diffraction patterns of 3d protein nano-crystals. *Zeitschrift für Kristallographie International journal for structural, physical, and chemical aspects of crystalline materials*, 225(2-3):67–76.
- [Achstetter and Wolf, 1985] Achstetter, T. and Wolf, D. H. (1985). Proteinases, proteolysis and biological control in the yeast *saccharomyces cerevisiae*. *Yeast*, 1(2):139–157.
- [Aoyama, 1996] Aoyama, K. (1996). 2d-crystallization of rhodococcus 20s proteasome at the liquid-liquid interface. *Journal of Crystal Growth*, 168(1–4):198–203.
- [Arfin and Bradshaw, 1988] Arfin, S. M. and Bradshaw, R. A. (1988). Cotranslational processing and protein turnover in eukaryotic cells. *Biochemistry*, 27(21):7979–7984.
- [Arndt and Wonacott, 1977] Arndt, W. and Wonacott, A., editors (1977). The rotation method in crystallography. *Amsterdam a.o. : North-Holland*.
- [ASI, 2015] ASI (2015). Amsterdam scientific instruments, science park 105 1098 xg amsterdam the netherlands.
- [Baldock et al., 1996] Baldock, P., Mills, V., and Stewart, P. S. (1996). A comparison of microbatch and vapour diffusion for initial screening of crystallization conditions. *Journal of crystal growth*, 168(1):170–174.
- [Baldwin et al., 1988] Baldwin, J. M., Henderson, R., Beckman, E., and Zemlin, F. (1988). Images of purple membrane at 2,8 Å resolution obtained by cryoelectron microscopy. *J Mol Biol*, 202(3):585–591.
- [Ballabriga et al., 2011] Ballabriga, R., Campbell, M., Heijne, E., Llopart, X., Tlustos, L., and Wong, W. (2011). Medipix3: A 64 k pixel detector readout chip working in single photon counting mode with improved spectrometric performance. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators; Spectrometers; Detectors and Associated Equipment*, 633, Supplement 1:S15 – S18. 11th International Workshop on Radiation Imaging Detectors (IWORID).
- [Bard et al., 2004] Bard, J., Ercolani, K., Svenson, K., Olland, A., and Somers, W. (2004). Automated systems for protein crystallization. *Methods*, 34(3):329–347.
- [Barrett and Kirschke, 1981] Barrett, A. J. and Kirschke, H. (1981). [41] cathepsin b, cathepsin h, and cathepsin l. In *Proteolytic Enzymes*, Part C, volume 80 of *Methods in Enzymology*, pages 535 – 561. Academic Press.
- [Bartesaghi et al., 2015] Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J. L. S., and Subramaniam, S. (2015). 2.2 Å resolution cryo-em structure of l2 – galactosidase in complex with a cell – permeant inhibitor. *Science*, 348(6239) : 1147 – 1151.
- [Batson et al., 2002] Batson, P. E., Dellby, N., and Krivanek, O. L. (2002). Sub-angstrom resolution using aberration corrected electron optics. *Nature*, 418(6898):617–620.
- [Bello et al., 2003] Bello, D. S. S., van Beuzekom, M., Jansweijer, P., Verkooijen, H., and Visschers, J. (2003). An interface board for the control and data acquisition of the medipix2 chip. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators; Spectrometers; Detectors and Associated Equipment*, 509(1–3):164 – 170. Proceedings of the 4th International Workshop on Radiation Imaging Detectors.
- [Bergfors, 2003] Bergfors, T. (2003). Seeds to crystals. *Journal of structural biology*, 142(1):66–76.
- [Bergfors, 2009] Bergfors, T. M. (2009). Protein crystallization. *Internat'l University Line*.
- [Berne and Pecora, 1976] Berne, B. and Pecora, R. (1976). Dynamic light scattering with applications to biology, chemistry and physics. *New York: Wiley*, 1:1.

## References

- [Berry et al., 2006] Berry, I. M., Dym, O., Esnouf, R., Harlos, K., Meged, R., Perrakis, A., Sussman, J., Walter, T., Wilson, J., and Messerschmidt, A. (2006). Spine high-throughput crystallization, crystal imaging and recognition techniques: current state, performance analysis, new technologies and future aspects. *Acta Crystallographica Section D: Biological Crystallography*, 62(10):1137–1149.
- [Bisogni et al., 1998] Bisogni, M. G. et al. (1998). Performance of a 4096 pixel photon counting chip. In *SPIE Proceedings*, 3445-31).
- [Bond and Butler, 1987] Bond, J. S. and Butler, P. E. (1987). Intracellular proteases. *Annual review of biochemistry*, 56(1):333–364.
- [Brönnimann et al., 2002] Brönnimann, C., Baur, R., Eikenberry, E., Fischer, P., Florin, S., Horisberger, R., Lindner, M., Schmitt, B., and Schulze, C. (2002). A pixel detector for the protein crystallography beamline at the {SLS}. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators; Spectrometers; Detectors and Associated Equipment*, 477(1–3):531–535. 5th Int. Conf. on Position-Sensitive Detectors.
- [Bruker, 2004] Bruker, A. (2004). Saint-plus and xprep. *Bruker Axs Inc.*, Madison, Wisconsin, USA.
- [Chapman et al., 2011] Chapman, H. N., Fromme, P., Barty, A., White, T. A., Kirian, R. A., Aquila, A., Hunter, M. S., Schulz, J., DePonte, D. P., Weierstall, U., Doak, R. B., Maia, F. R. N. C., Martin, A. V., Schlichting, I., Lomb, L., Coppola, N., Shoeman, R. L., Epp, S. W., Hartmann, R., Rolles, D., Rudenko, A., Foucar, L., Kimmel, N., Weidenspointner, G., Holl, P., Liang, M., Barthelmess, M., Caleman, C., Boutet, S., Bogan, M. J., Krzywinski, J., Bostedt, C., Bajt, S., Gumprecht, L., Rudek, B., Erk, B., Schmidt, C., Homke, A., Reich, C., Pietschner, D., Struder, L., Hauser, G., Gorke, H., Ullrich, J., Herrmann, S., Schaller, G., Schopper, F., Soltau, H., Kuhnelt, K. U., Messerschmidt, M., Bozek, J. D., Hau-Riege, S. P., Frank, M., Hampton, C. Y., Sierra, R. G., Starodub, D., Williams, G. J., Hajdu, J., Timneanu, N., Seibert, M. M., Andreasson, J., Rucker, A., Jonsson, O., Svenda, M., Stern, S., Nass, K., Andrichke, R., Schroter, C.-D., Krasniqi, F., Bott, M., Schmidt, K. E., Wang, X., Grotjohann, I., Holton, J. M., Barends, T. R. M., Neutze, R., Marchesini, S., Fromme, R., Schorb, S., Rupp, D., Adolph, M., Gorkhover, T., Andersson, I., Hirsemann, H., Potdevin, G., Graafsma, H., Nilsson, B., and Spence, J. C. H. (2011). Femtosecond x-ray protein nanocrystallography. *Nature*, 470(7332):73–77.
- [Cheng, 2013] Cheng, D. (2013). Field and Wave Electromagnetics. *Pearson Education*, Limited.
- [Cork et al., 2006] Cork, C., O'Neill, J., Taylor, J., and Earnest, T. (2006). Advanced beamline automation for biological crystallography experiments. *Acta Crystallographica Section D: Biological Crystallography*, 62(8):852–858.
- [Dalling, 1986] Dalling, M. J. (1986). Plant proteolytic enzymes. *CRC Press*. [Eddleston et al., 2010] Eddleston, M. D., Bithell, E. G., and Jones, W. (2010). Transmission electron microscopy of pharmaceutical materials. *J. Pharm. Sci.*, 99:4072–4083.
- [El Hassan et al., 2013] El Hassan, N., Ikni, A., Gillet, J.-M., Spasojevic-de Biré, A., and Ghermani, N. E. (2013). Electron properties of carbamazepine drug in form iii. *Crystal Growth & Design*, 13:2887–2896.
- [Evans, 2006] Evans, P. (2006). Scaling and assessment of data quality. *Acta Cryst. D*, 62(1):72–82.
- [Faruqi and Henderson, 2007] Faruqi, A. R. and Henderson, R. (2007). Electronic detectors for electron microscopy. *Curr Opin Struct Biol*, 17(5):549–555.
- [Faruqi and McMullan, 2011] Faruqi, A. R. and McMullan, G. (2011). Electronic detectors for electron microscopy. *Quarterly Reviews of Biophysics*, 44:357–390.
- [FEI, 2015] FEI (2015). Fei, 5350 ne, dawson creek drive hillsboro, oregon 97124, usa.
- [Gatan, 2015] Gatan (2015). Gatan inc., 5794 w. las positas blvd. pleasanton, ca 94588, united states.

- [**Gemmi and Nicolopoulos, 2007**] Gemmi, M. and Nicolopoulos, S. (2007). Structure solution with three-dimensional sets of precessed electron diffraction intensities. *Ultramicroscopy*, 107:483–494. Proc. Electron Cryst. School 2005, {ELCRYST} 2005: New Frontiers in Electron Crystallography.
- [**Georgieva et al., 2011**] Georgieva, D., Jansen, J., Sikharulidze, I., Jiang, L., Zandbergen, H. W., and Abrahams, J. P. (2011). Evaluation of medipix2 detector for recording electron diffraction data in low dose conditions. *Journal of Instrumentation*, 6(01):C01033.
- [**Giegé, 1987**] Giegé, R. (1987). Crystallization of protein and nucleic acids: A survey of methods and importance of the purity of the macromolecules. In *Crystallography in Molecular Biology*, pages 15–26. Springer.
- [**Gil et al., 2006**] Gil, D., Carazo, J. M., and Marabini, R. (2006). On the nature of 2d crystal unbending. *Journal of Structural Biology*, 156(3):546 – 555.
- [**Gorelik et al., 2012**] Gorelik, T. E., van de Streek, J., Kilbinger, A. F. M., Brunklaus, G., and Kolb, U. (2012). Ab-initio crystal structure analysis and refinement approaches of oligo p-benzamides based on electron diffraction data. *Acta Cryst. B*, 68:171–181.
- [**Grune, 2008**] Grune, T. (2008). mtz2sca and mtz2hkl : facilitated transition from CCP4 to the SHELX program suite. *Journal of Applied Crystallography*, 41(1):217–218.
- [**Hamann et al., 2015**] Hamann, E., Koenig, T., Zuber, M., Cecilia, A., Tyazhev, A., Tolbanov, O., Procz, S., Fauler, A., Baumbach, T., and Fiederle, M. (2015). Performance of a medipix3rx spectroscopic pixel detector with a high resistivity gallium arsenide sensor. *Medical Imaging, IEEE Transactions on*, 34(3):707–715.
- [**Hattne et al., 2015**] Hattne, J., Reyes, F. E., Nannenga, B. L., Shi, D., de la Cruz, M. J., Leslie, A. G. W., and Gonen, T. (2015). MicroED data collection and processing. *Acta Crystallographica Section A*, 71(4):353–360.
- [**Henderson and Unwin, 1975**] Henderson, R. and Unwin, P. N. T. (1975). Three-dimensional model of purple membrane obtained by electron microscopy. *Nature*, 257(5521):28–32.
- [**Henderson et al., 1990**] Henderson, R., Baldwin, J., Ceska, T., Zemlin, F., Beckmann, E., and Downing, K. (1990). Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *Journal of Molecular Biology*, 213(4):899 – 929.
- [**Henderson, 1995**] Henderson, R. (1995). The potential and limitations of neutrons, electrons and x-rays for atomic resolution microscopy of unstained biological molecules. *Q Rev Biophys*, 28(2):171–193.
- [**Hübschle et al., 2011**] Huübschle, C. B., Sheldrick, G. M., and Dittrich, B. (2011). ShelXle: a Qt graphical user interface for SHELXL. *Journal of Applied Crystallography*, 44:1281–1284.
- [**Jansen et al., 1998**] Jansen, J., Tang, D., Zandbergen, H. W., and Schenk, H. (1998). MSLS, a Least-Squares Procedure for Accurate Crystal Structure Refinement from Dynamical Electron Diffraction Patterns. *Acta Crystallographica Section A*, 54(1):91–101.
- [**Jiang et al., 2009**] Jiang, L., Georgieva, D., Zandbergen, H. W., and Abrahams, J. P. (2009). Unit-cell determination from randomly oriented electron diffraction patterns. *Acta Crystallographica Section D*, 65(7):625–632.
- [**Jiang et al., 2011**] Jiang, L., Georgieva, D., Nederlof, I., Liu, Z., and Abrahams, J. P. (2011). Image processing and lattice determination for threedimensional nanocrystals. *Microscopy and Microanalysis*, 17:879–885.
- [**Judge et al., 2005**] Judge, R. A., Swift, K., and Gonzalez, C. (2005). An ultraviolet fluorescence-based method for identifying and distinguishing protein crystals. *Acta Crystallographica Section D: Biological Crystallography*, 61(1):60–66.
- [**Kabsch, 2010**] Kabsch, W. (2010). XDS. *Acta Cryst. D*, 66:125–132.



## References

- [Kang et al., 2015] Kang, Y., Zhou, X. E., Gao, X., He, Y., Liu, W., Ishchenko, A., Barty, A., White, T. A., Yefanov, O., Han, G. W., Xu, Q., de Waal, P. W., Ke, J., Tan, M. H. E., Zhang, C., Moeller, A., West, G. M., Pascal, B. D., Van Eps, N., Caro, L. N., Vishnivetskiy, S. A., Lee, R. J., SuinoPowell, K. M., Gu, X., Pal, K., Ma, J., Zhi, X., Boutet, S., Williams, G. J., Messerschmidt, M., Gati, C., Zatsepin, N. A., Wang, D., James, D., Basu, S., Roy-Chowdhury, S., Conrad, C. E., Coe, J., Liu, H., Lisova, S., Kupitz, C., Grotjohann, I., Fromme, R., Jiang, Y., Tan, M., Yang, H., Li, J., Wang, M., Zheng, Z., Li, D., Howe, N., Zhao, Y., Standfuss, J., Diederichs, K., Dong, Y., Potter, C. S., Carragher, B., Caffrey, M., Jiang, H., Chapman, H. N., Spence, J. C. H., Fromme, P., Weierstall, U., Ernst, O. P., Katritch, V., Gurevich, V. V., Griffin, P. R., Hubbell, W. L., Stevens, R. C., Cherezov, V., Melcher, K., and Xu, H. E. (2015). Crystal structure of rhodopsin bound to arrestin by femtosecond x-ray laser. *Nature*, 523(7562):561–567.
- [Karger and Hancock, 1996] Karger, B. L. and Hancock, W. S. (1996). High Resolution Separation and Analysis of Biological Macromolecules, volume 270. *Elsevier*.
- [Karplus and Diederichs, 2012] Karplus, P. A. and Diederichs, K. (2012). Linking crystallographic model and data quality. *Science*, 336:1030–1033.
- [Kirian et al., 2011] Kirian, R. A., White, T. A., Holton, J. M., Chapman, H. N., Fromme, P., Barty, A., Lomb, L., Aquila, A., Maia, F. R. N. C., Martin, A. V., Fromme, R., Wang, X., Hunter, M. S., Schmidt, K. E., and Spence, J. C. H. (2011). Structure-factor analysis of femtosecond microdiffraction patterns from protein nanocrystals. *Acta Crystallographica Section A: Foundations of Crystallography*, 67(Pt 2):131–140.
- [Klug, 1979] Klug, A. (1979). Image analysis and reconstruction in the electron microscopy of biological macromolecules. *Chemica Scripta*, 14:245–256.
- [Kolb et al., 2007] Kolb, U., Gorelik, T., Kuübel, C., Otten, M., and Hubert, D. (2007). Towards automated diffraction tomography: Part i—data acquisition. *Ultramicroscopy*, 107(6–7):507 – 513. Proceedings of the Electron Crystallography School 2005, {ELCRYST} 2005: New Frontiers in Electron Crystallography.
- [Kolb et al., 2008] Kolb, U., Gorelik, T., and Otten, M. (2008). Towards automated diffraction tomography; part ii—cell parameter determination. *Ultramicroscopy*, 108(8):763 – 772.
- [Kolb et al., 2010] Kolb, U., Gorelik, T. E., Mugnaioli, E., and Stewart, A. (2010). Structural characterization of organics using manual and automated electron diffraction. *Polym. Rev.*, 50:385–409.
- [Kolb et al., 2011] Kolb, U., Mugnaioli, E., and Gorelik, T. E. (2011). Automated electron diffraction tomography – a new tool for nano crystal structure analysis. *Cryst. Res. Technol.*, 46:542–554.
- [Konev, 1967] Konev, S. V. (1967). Electronic excited states of proteins. *Springer*.
- [Kraus et al., 2011] Kraus, V., Holik, M., Jakubek, J., Kroupa, M., Soukup, P., and Vykydal, Z. (2011). Fitpix — fast interface for timepix pixel detectors. *Journal of Instrumentation*, 6(01):C01079.
- [Kuhlbrandt and Wang, 1991] Kuhlbrandt, W. and Wang, D. N. (1991). Three-dimensional structure of plant light-harvesting complex determined by electron crystallography. *Nature*, 350(6314):130–134.
- [Kutoglu and Scheringer, 1983] Kutoglu, A. and Scheringer, C. (1983). Nicotinic acid, C<sub>6</sub>H<sub>5</sub>NO<sub>2</sub>: refinement. *Acta Cryst. C*, 39:232–234.
- [Le Maire et al., 2011] Le Maire, A., Gelin, M., Pochet, S., Hoh, F., Pirocchi, M., Guichou, J.-F., Ferrer, J.-L., and Labesse, G. (2011). In-plate protein crystallization, in situ ligand soaking and x-ray diffraction. *Acta Crystallographica Section D: Biological Crystallography*, 67(9):747–755.
- [Llopart et al., 2002] Llopart, X., Campbell, M., Dinapoli, R., San Segundo, D., and Pernigotti, E. (2002). Medipix2: A 64-k pixel readout chip with 55μm square elements working in single photon counting mode. *Nuclear Science, IEEE Transactions on*, 49(5):2279–2283.



- [**Llopart et al., 2007**] Llopart, X., Ballabriga, R., Campbell, M., Tlustos, L., and Wong, W. (2007). Timepix, a 65k programmable pixel readout chip for arrival time, energy and/or photon counting measurements. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators; Spectrometers; Detectors and Associated Equipment*, 581(1–2):485 – 494. {VCI} 2007 Proceedings of the 11th International Vienna Conference on Instrumentation.
- [**Lomb et al., 2011**] Lomb, L., Barends, T. R. M., Kassemeyer, S., Aquila, A., Epp, S. W., Erk, B., Foucar, L., Hartmann, R., Rudek, B., Rolles, D., Rudenko, A., Shoeman, R. L., Andreasson, J., Bajt, S., Barthelmess, M., Barty, A., Bogan, M. J., Bostedt, C., Bozek, J. D., Caleman, C., Coffee, R., Coppola, N., DePonte, D. P., Doak, R. B., Ekeberg, T., Fleckenstein, H., Fromme, P., Gebhardt, M., Graafsma, H., Gumprecht, L., Hampton, C. Y., Hartmann, A., Hauser, G., Hirsemann, H., Holl, P., Holton, J. M., Hunter, M. S., Kabsch, W., Kimmel, N., Kirian, R. A., Liang, M., Maia, F. R. N. C., Meinhart, A., Marchesini, S., Martin, A. V., Nass, K., Reich, C., Schulz, J., Seibert, M. M., Sierra, R., Soltau, H., Spence, J. C. H., Steinbrener, J., Stellato, F., Stern, S., Timneanu, N., Wang, X., Weidenspointner, G., Weierstall, U., White, T. A., Wunderer, C., Chapman, H. N., Ullrich, J., Strüder, L., and Schlichting, I. (2011). Radiation damage in protein serial femtosecond crystallography using an x-ray free-electron laser. *Phys. Rev. B*, 84:214111.
- [**Luebben and Gruene, 2015**] Luebben, J. and Gruene, T. (2015). New method to compute Rcomplete enables maximum likelihood refinement for small data sets. *Proc. Natl. Acad. Sci. U. S. A.*, 112:8999–9003.
- [**Marko and Danev, 2012**] Marko, M. and Danev, R. (2012). A brief introduction to phase plates: Benefits and types. *Microscopy and Microanalysis*, 18:460–461.
- [**Matthews, 1968**] Matthews, B. W. (1968). Solvent content of protein crystals. *Journal of molecular biology*, 33(2):491–497.
- [**McMullan et al., 2007**] McMullan, G., Cattermole, D., Chen, S., Henderson, R., Llopart, X., Summerfield, C., Tlustos, L., and Faruqi, A. (2007). Electron imaging with medipix2 hybrid pixel detector. *Ultramicroscopy*, 107(4–5):403– 413.
- [**McMullan et al., 2009**] McMullan, G., Chen, S., Henderson, R., and Faruqi, A. R. (2009). Detective quantum efficiency of electron area detectors in electron microscopy. *Ultramicroscopy*, 109:1126–1143.
- [**McMullan et al., 2014**] McMullan, G., Faruqi, A., Clare, D., and Henderson, R. (2014). Comparison of optimal performance at 300 kev of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy*, 147:156 – 163.
- [**McPherson and Shlichta, 1988**] McPherson, A. and Shlichta, P. (1988). The use of heterogeneous and epitaxial nucleants to promote the growth of protein crystals. *Journal of Crystal Growth*, 90(1):47–50.
- [**McPherson et al., 1996**] McPherson, A., Malkin, A. J., Kuznetsov, Y. G., and Koszelak, S. (1996). Incorporation of impurities into macromolecular crystals. *Journal of crystal growth*, 168(1):74–92.
- [**McPherson, 1985**] McPherson, A. (1985). [5] crystallization of macromolecules: General principles. *Methods in enzymology*, 114:112–120.
- [**Mikol et al., 1989**] Mikol, V., Hirsch, E., and Giegé, R. (1989). Monitoring protein crystallization by dynamic light scattering. *FEBS letters*, 258(1):63– 66.
- [**Mikol et al., 1990**] Mikol, V., Hirsch, E., and Giegé, R. (1990). Diagnostic of precipitant for biomacromolecule crystallization by quasi-elastic lightscattering. *Journal of molecular biology*, 213(1):187–195.
- [**Mugnaioli and Kolb, 2014**] Mugnaioli, E. and Kolb, U. (2014). Structure solution of zeolites by automated electron diffraction tomography – impact and treatment of preferential orientation. *Microporous and Mesoporous Materials*, 189:107–114.
- [**Mugnaioli et al., 2009**] Mugnaioli, E., Gorelik, T. E., and Kolb, U. (2009). “ab initio” structure solution from electron diffraction data obtained by a combination of automated diffraction tomography and precession technique. *Ultramicroscopy*, 109:758–765.
-

## References

---

- [Nederlof et al., 2011] Nederlof, I., Hosseini, R., Georgieva, D., Luo, J., Li, D., and Abrahams, J. P. (2011). A straightforward and robust method for introducing human hair as a nucleant into high throughput crystallization trials. *Crystal Growth & Design*, 11(4):1170–1176.
- [Nederlof et al., 2013a] Nederlof, I., Li, Y. W., van Heel, M., and Abrahams, J. P. (2013a). Imaging protein three-dimensional nanocrystals with cryo-EM. *Acta Crystallographica Section D*, 69(5):852–859.
- [Nederlof et al., 2013b] Nederlof, I., van Genderen, E., Li, Y.-W., and Abrahams, J. P. (2013b). A Medipix quantum area detector allows rotation electron diffraction data collection from submicrometre three-dimensional protein crystals. *Acta Crystallographica Section D*, 69(7):1223–1230.
- [Ozaki et al., 2014] Ozaki, S., Nakagawa, Y., Shirai, O., and Kano, K. (2014). Substituent effect on the thermodynamic solubility of structural analogs: Relative contribution of crystal packing and hydration. *Journal of Pharmaceutical Sciences*, 103(11):3524–3531.
- [Peng, 1999] Peng, L.-M. (1999). Electron atomic scattering factors and scattering potentials of crystals. *Micron*, 30:625–648.
- [Perrin and Dempsey, 1974] Perrin, D. and Dempsey, B. (1974). Buffers for pH and metal ion control.
- [Plackett et al., 2010] Plackett, R., Ballabriga, R., Campbell, M., Llopart, X., Tick, T., Tlustos, L., Turecek, D., Vahnen, S., and Wong, W. (2010). Medipix 3 and timepix 2 chip and applications. In *PoS (VERTEX 2010) 030*.
- [Press et al., 2007] Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). Numerical Recipes: the art of scientific computing, Third Edition (C++), volume 994. *Cambridge University Press*.
- [Rasif, 2015] Rasif, M. Z. (2015). CdTe/CdZnTe pixellated radiation detector. *PhD thesis*, University of Glasgow.
- [Raunser and Walz, 2009] Raunser, S. and Walz, T. (2009). Electron crystallography as a technique to study the structure on membrane proteins in a lipidic environment. *Annual Review of Biophysics*, 38(1):89–105. PMID: 19416061.
- [Robert and Lefauchaux, 1988] Robert, M. and Lefauchaux, F. (1988). Crystal growth in gels: principle and applications. *Journal of Crystal Growth*, 90(1):358–367.
- [Rose, 2009] Rose, H. H. (2009). Future trends in aberration-corrected electron microscopy. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 367(1903):3809–3823.
- [Rupp, 2004] Rupp, B. (2004). A guide to automation and data handling in protein crystallization, Edited by N Chayen. *Protein Crystallization Strategies for Structural Genomics. International University Line*.
- [Salemme, 1972] Salemme, F. (1972). A free interface diffusion technique for the crystallization of proteins for x-ray crystallography. *Archives of Biochemistry and Biophysics*, 151(2):533–539.
- [Scherer et al., 2014] Scherer, S., Arheit, M., Kowal, J., Zeng, X., and Stahlberg, H. (2014). Single particle 3d reconstruction for 2d crystal images of membrane proteins. *Journal of Structural Biology*, 185(3):267 – 277.
- [Shaw, 1978] Shaw, R. (1978). Evaluating the efficient of imaging processes. *Reports on Progress in Physics*, 41(7):1103.
- [Sheldrick, 2008] Sheldrick, G. M. (2008). A short history of SHELX . *Acta Cryst.*, A64:112–122.
- [Sheldrick, 2015] Sheldrick, G. M. (2015). SHELXT – Integrated space-group and crystal-structure determination. *Acta Cryst.*, A71:3–8.
- [Skarzynski, 2009] Skarzynski, T. (2009). X-ray diffraction from crystals in crystallization plates. *Acta Crystallographica Section A*, 65(a1):s159.
-

- [Skouri et al., 1995] Skouri, M., Lorber, B., Giegé, R., Munch, J.-P., and Candau, J. (1995). Effect of macromolecular impurities on lysozyme solubility and crystallizability: dynamic light scattering, phase diagram, and crystal growth studies. *Journal of Crystal Growth*, 152(3):209 – 220.
- [Song et al., 2007] Song, J., Mathew, D., Jacob, S. A., Corbett, L., Moorhead, P., and Soltis, S. M. (2007). Diffraction-based automated crystal centering. *Journal of synchrotron radiation*, 14(2):191–195.
- [Stahlberg et al., 2001] Stahlberg, H., Fotiadis, D., Scheuring, S., Rémigy, H., Braun, T., Mitsuoaka, K., Fujiyoshi, Y., and Engel, A. (2001). Twodimensional crystals: a powerful approach to assess structure, function and dynamics of membrane proteins. *{FEBS} Letters*, 504(3):166 – 172. Structure, Dynamics and Function of Proteins in Biological Membranes.
- [Stevens, 2000] Stevens, R. C. (2000). High-throughput protein crystallization. *Current opinion in structural biology*, 10(5):558–563.
- [Sugahara et al., 2008] Sugahara, M., Asada, Y., Morikawa, Y., Kageyama, Y., and Kunishima, N. (2008). Nucleant-mediated protein crystallization with the application of microporous synthetic zeolites. *Acta Crystallogr D Biol Crystallogr*, 64(Pt 6):686–695.
- [Takehara et al., 2008] Takehara, M., Ino, K., Takakusagi, Y., Oshikane, H., Nureki, O., Ebina, T., Mizukami, F., and Sakaguchi, K. (2008). Use of layer silicate for protein crystallization: Effects of micromica and chlorite powders in hanging drops. *Analytical biochemistry*, 373(2):322–329.
- [Taylor and Glaeser, 1974] Taylor, K. and Glaeser, R. (1974). Electron diffraction of frozen, hydrated protein crystals. *Science*, 186(4168):1036–7.
- [Thorn et al., 2012] Thorn, A., Dittrich, B., and Sheldrick, G. M. (2012). Enhanced rigid-bond restraints. *Acta Cryst*, A68:448–451.
- [Turecek et al., 2011] Turecek, D., Holy, T., Jakubek, J., Pospisil, S., and Vykydal, Z. (2011). Pixelman: a multi-platform data acquisition and processing software package for medipix2, timepix and medipix3 detectors. *Journal of Instrumentation*, 6(01):C01046.
- [Van der Laan et al., 1989] Van der Laan, J. M., Swarte, M. B., Groendijk, H., Hol, W. G., and Drenth, J. (1989). The influence of purification and protein heterogeneity on the crystallization of p-hydroxybenzoate hydroxylase. *Eur J Biochem*, 179(3):715–724.
- [Vekilov and Rosenberger, 1996] Vekilov, P. and Rosenberger, F. (1996). Dependence of lysozyme growth kinetics on step sources and impurities. *Journal of Crystal Growth*, 158(4):540 – 551.
- [Visser et al., 2011] Visser, J., van der Heijden, B., Weijers, S., de Vries, R., and Visschers, J. (2011). A gigabit per second read-out system for medipix quads. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators; Spectrometers; Detectors and Associated Equipment*, 633, Supplement 1:S22 – S25. 11th International Workshop on Radiation Imaging Detectors (IWORID).
- [Vykydal et al., 2006] Vykydal, Z., Jakubek, J., and Pospisil, S. (2006). Usb interface for medipix2 pixel device enabling energy and position-sensitive detection of heavy charged particles. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators; Spectrometers; Detectors and Associated Equipment*, 563(1):112–115.
- [Wampler et al., 2008] Wampler, R. D., Kissick, D. J., Dehen, C. J., Gualtieri, E. J., Grey, J. L., Wang, H.-F., Thompson, D. H., Cheng, J.-X., and Simpson, G. J. (2008). Selective detection of protein crystals by second harmonic microscopy. *J Am Chem Soc*, 130(43):14076–14077.
- [Wan et al., 2013] Wan, W., Sun, J., Su, J., Hovmöller, S., and Zou, X. (2013). Three-dimensional rotation electron diffraction: software RED for automated data collection and data processing. *J Appl. Cryst.*, 46:1863–1873.
- [Wandersman, 1989] Wandersman, C. (1989). Secretion, processing and activation of bacterial extracellular proteases. *Mol Microbiol*, 3(12):1825–1831.
- [Wang, 1995] Wang, Z. (1995). Elastic and Inelastic Scattering in Electron Diffraction and Imaging. Plenum Press.

## References

---

- [Wierenga et al., 1987] Wierenga, R. K., Kalk, K. H., and Hol, W. G. (1987). Structure determination of the glycosomal triosephosphate isomerase from trypanosoma brucei brucei at 2.4 Å resolution. *J Mol Biol*, 198(1):109–121.
- [Williams and Barry Carter, 2009] Williams, D. B. and Barry Carter, C. (2009). Transmission Electron Microscopy A Textbook for Materials Science. Number SBN 978-0-387-76501-3 in 2nd Edition. Springer US.
- [Winn et al., 2011] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011). Overview of the CCP 4 suite and current developments. *Acta Cryst. D*, 67:235–242.
- [Wüthrich, 1995] Wüthrich, K. (1995). NMR – this other method for protein and nucleic acid structure determination. *Acta Crystallographica Section D*, 51(3):249–270.
- [Yun et al., 2015] Yun, Y., Zou, X., Hovmöller, S., and Wan, W. (2015). Three-dimensional electron diffraction as a complementary technique to powder X-ray diffraction for phase identification and structure solution of powders. *IUCrJ*, 2:267–282.
- [Zeng et al., 2012] Zeng, X., Chen, Y.-W., Hughes, O., and Stahlberg, H. (2012). Iterative transform algorithms for 3d reconstruction of 2d crystals. *JNIT: Journal of Next Generation Information Technology*, 3(1):28–35.
- [Zhang et al., 2010] Zhang, D., Oleynikov, P., Hovmöller, S., and Zou, X. (2010). Collecting 3d electron diffraction data by the rotation method. *Z. Kristallogr. – Cryst. Mater.*, 225:94.
- [Zwerger et al., 2007] Zwerger, A., Fauler, A., Fiederle, M., and Jakobs, K. (2007). Medipix2: Processing and measurements of gas pixel detectors. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators; Spectrometers; Detectors and Associated Equipment*, 576(1):23–26.

## List of Abbreviations

2D	—	Two-dimensional
3D	—	Three-dimensional
ADT	—	automated diffraction tomography
ASI	—	Amsterdam Scientific Instruments
ASIC	—	Application-specific integrated circuit
ATP	—	Adenosine triphosphate
CCD	—	Charge-coupled device
CdTe	—	Cadmium telluride
Cl	—	Chlorite
CM	—	Counting mode
CTF	—	Contrast transfer function
DAC	—	Data acquisition control
Disc	—	Discriminator
DQE	—	Data quantum efficiency
DT	—	Dead time
ED	—	Electron diffraction
EG	—	Electron Gun
EM	—	Electron microscopy
FEG	—	Field emission gun
FEL	—	Free electron laser
GaAs	—	Gallium arsenide
HRTEM	—	High resolution transmission electron microscop(y/e)
HVEM	—	High-voltage electron microscop(y/e)
MPD	—	2-Methyl-2,4-pentanediol
MTF	—	Modulation transfer function
Na	—	Sodium
NPS	—	Noise Power Spectrum
PEG	—	Polyethylene glycol
PSF	—	Point spread function
Quad	—	Quadruple
RED	—	rotation electron diffraction
RT	—	Room temperature
RVS	—	Stainless steel
S/N	—	Signal to noise
Sb	—	Antimony
SHG	—	Second harmonic generation

---

Si	—	Silicon
SONICC	—	Second harmonic microscopy
SSNMR	—	Solid state nuclear magnetic resonance
STEM	—	Scanning transmission electron microscop(y/e)
TEM	—	Transmission electron microscop(y/e)
THL	—	Threshold lower
TOA	—	Time of arrival
TOT	—	Time over threshold
UV	—	Ultra-violet
XRPD	—	X-ray powder diffraction
Z	—	Atomic number

## List of Figures

<b>Figure 1.1:</b> Typical transmission electron microscope schematic. ....	<b>11</b>
<b>Figure 1.2:</b> Scattering in crystals. ....	<b>13</b>
<b>Figure 1.3:</b> Chemical structure perylene derivative. ....	<b>14</b>
<b>Figure 1.4:</b> Diffraction pattern from perylene derivative. ....	<b>15</b>
<b>Figure 2.1:</b> Counting modes in Timepix chip. ....	<b>27</b>
<b>Figure 2.2:</b> Monte Carlo simulations of electron scattering. ....	<b>27</b>
<b>Figure 2.3:</b> Simplified schematic of the counting electronics of a Timepix chip. ....	<b>31</b>
<b>Figure 2.4:</b> Cooling test set-up and results. ....	<b>33</b>
<b>Figure 2.5:</b> 3D 90° cut of the YEOL 1010 camera housing. ....	<b>33</b>
<b>Figure 2.6:</b> 3D design drawings of the “Hardy” Krios camera. ....	<b>35</b>
<b>Figure 2.7:</b> Design schematic for: cable management, sensors, data transport and interlocks ....	<b>35</b>
<b>Figure 2.8:</b> 3D design drawings of the “Laurel” Juelich camera. ....	<b>39</b>
<b>Figure 3.1:</b> Crystal packing in lysozyme crystals. ....	<b>47</b>
<b>Figure 3.2:</b> Solubility curve of a protein. ....	<b>47</b>
<b>Figure 3.3:</b> Glucose isomerase at different crystallization conditions. ....	<b>51</b>
<b>Figure 3.4:</b> Schematic of different crystallization methods. ....	<b>51</b>
<b>Figure 3.5:</b> Lysozyme needle crystals growing on sliced human hair. ....	<b>51</b>
<b>Figure 4.1:</b> Lysozyme nano-crystals. ....	<b>67</b>
<b>Figure 4.2:</b> Cut-out view of the Medipix2 camera assembly. ....	<b>67</b>
<b>Figure 4.3:</b> Three diffraction patterns. ....	<b>71</b>
<b>Figure 4.4:</b> Overlay of predicted spot positions. ....	<b>71</b>
<b>Figure 4.5:</b> Spot profile from MOSFLM for the first data set. ....	<b>71</b>
<b>Figure 4.6:</b> Laue zones at $h=0$ of the scaled data. ....	<b>73</b>
<b>Figure 5.1:</b> Projection of $ <F_i(y)> $ onto $F_m(y)$ ....	<b>84</b>
<b>Figure 5.2:</b> Original images and filtered results. ....	<b>89</b>
<b>Figure 5.3:</b> Unique half of the centro-symmetric lattice filter of the lysozyme crystal image ....	<b>91</b>
<b>Figure 5.4:</b> Filtered random noise. ....	<b>91</b>
<b>Figure 5.5:</b> Plot of the rotational average. ....	<b>91</b>
<b>Figure 5.6:</b> Filtering results from extreme variables. ....	<b>92</b>
<b>Figure 6.1:</b> “Laurel” Timepix detector. ....	<b>101</b>
<b>Figure 6.2:</b> Timepix detector quality. ....	<b>101</b>
<b>Figure 6.3:</b> Rotation electron diffraction of pharmaceutical nanocrystals. ....	<b>105</b>
<b>Figure 6.4:</b> Unit cell content for the final refined structures. ....	<b>105</b>
<b>Figure 6.S1:</b> How a short wavelength leads to greater errors in unit cell parameters. ....	<b>112</b>

---

## List of Tables

<b>Table 1.1:</b> Energy deposited in biological/organic specimens per useful scattering event. ....	12
<b>Table 2.1:</b> Comparison table between currently available chips. ....	24
<b>Table 3.1:</b> Parameters influencing the crystallization of biomacromolecules. ....	44
<b>Table 6.1:</b> Data statistics from XDS and model R-values after refinement with SHELXL. ....	109
<b>Table 6.S1:</b> Carbamasepine, merged data from five crystals. ....	114
<b>Table 6.S2:</b> Nicotinic Acid, merged data from five crystals. ....	115
<b>Table 6.S3:</b> Bond length for carbamazepine (carb.) compared to the high resolution structure. ....	116
<b>Table 6.S4:</b> Selected bond angles for carb. compared to the high resolution structure.....	117
<b>Table 6.S5:</b> Bond length for nicotinic acid (nic.) compared to the high resolution structure.....	117
<b>Table 6.S6:</b> Selected bond angles for nic. compared to the high resolution structure .....	117



## Publications

van Genderen, E., Clabbers M.T.B., Pratim Das, P., Stewart, A., Nederlof, I., Barentsen, K.C., Portillo, J., Pannu, N.S., Nicolopoulos, S., Gruene, T., and Abrahams, J.P.

*Ab initio structure determination of nanocrystals of organic pharmaceutical compounds by electron diffraction at room temperature using a Timepix quantum area direct electron detector*

**(Submitted)**

van Genderen, E., Li Y-W, Nederlof, I., Abrahams, J.P.

*Lattice filter for processing image data of 3D protein nano-crystals.*

**(Submitted)**

Nederlof, I., van Genderen, E., Li, Y.W., Abrahams, J.P. **(2013).**

*A Medipix quantum area detector allows rotation electron diffraction data collection from submicrometre three-dimensional protein crystals*

**Acta Crystallographica Section D-Biological Crystallography**, vol. 69, pp. 1223-1230.

Nederlof, I., van Genderen, E., Hoedemaeker, F., Abrahams, J.P., Georgieva, D. **(2013).**

*Protein Crystal Growth.*

**Modern Aspects of Bulk Crystal and Thin Film Preparation: InTech.**

Li, N., Kuo, C.L., Paniagua, G., van den Elst, H., Verdoes, M., Willems, L. I., van der Linden, W.A., Ruben, M., van Genderen, E., Gubbens, J., van Wezel, G.P., Overkleeft, H.S., Florea, B. I **(2013)**

*Relative quantification of proteasome activity by activity-based protein profiling and LC-MS/MS,*

**Nature Protocols**, vol. 8, no. 6, pp. 1155-68.

---

## Curriculum Vitae

Eric van Genderen was born in Delft, the Netherlands on the 24th of December 1981. He went to high school at the Montessori Lyceum Rotterdam where he graduated in 2000. During this same year he joined the Twente University, where he started the undergraduate track electrical engineering. In 2002 he switched to the University of Leiden where he continued his studies in Life Science & Technology. Completing his internship at the Bio-organic chemistry group under supervision of Dr. B. Florea and Prof. Dr. H. Overkleef with a thesis on the topic "In vivo chemical capture of the Proteasome and identification by the use of an Mass Spectrometer (MS/MS)". He received his BSc. degree in 2008. He continued his undergraduate with a master in Functional Genomics - Life Science and Technology. His first internship was as Marketing Assistant at Cell Signaling Technology B.V., Leiden, the Netherlands under supervision of K. van den Berg, Prof. Dr. J. den Dunnen and Prof. Dr. J. Brouwer as his tutor at the University of Leiden. His second internship was at the Medical Hospital (LUMC), Leiden, the Netherlands, under supervision of Prof. Dr. J. den Dunnen and Dr. Y. Ariyurek, where he developed methods to use synthesised oligos, which were cleaved from a glass slide, for mass screening of selected DNA strings with different sequencing applications.

After he obtained his Masters degree in Januari 2010 he started May the same year his PhD at the University of Leiden at the Chemistry Institute in the Biophysical Structural Chemistry Group, headed by Prof. Dr. J.P. Abrahams. His research was focused on improvements of structure reconstruction of 3D nano-crystals from organic compounds with transmission electron microscopy. The main focus was on improving the detection methods for electrons by developing - and supervising the development - of novel detectors. Most of his work is presented in this thesis and he finished his PhD in 2015. His research was funded and supported by the University of Leiden, NWO, STW, Cyttron I, Cyttron II, Medipix consortium and Nanomegas company.

During his PhD he presented his work on multiple international conferences. Also he was involved as spokesman from the University of Leiden during 'closed' meetings of the Medipix consortium for developing new detectors of the 3th generation.

He is currently employed as a Postdoc at the University of Basel, Switzerland, at the Biozentrum in the C-CINA group. There, he will continue his research as described in this thesis, with an emphasis on developing electron diffraction technologies. The research is a joint project of the University of Basel, Switzerland, and the Paul Sherrer Institute, Switzerland.

